MEASUREMENT OF BIOMASS CONCENTRATION USING A MICROWAVE OVEN AND ANALYSIS OF DATA FOR ESTIMATION OF SPECIFIC RATES

by

MARK ANTHONY BUONO

B.S., Lehigh University, 1983

A MASTER'S THESIS

submitted in partial fulfillment of the

requirements for the degree

MASTER OF SCIENCE

CHEMICAL ENGINEERRING

Department of Chemical Engineering

KANSAS STATE UNIVERSITY

Manhattan, Kansas

1985

Approved by:

Major Professor

LD
2668
.T5
1985
B86
c.2

A11202 942461

PREFACE

The scope of this work is two fold. The first portion of this thesis addresses the task of measuring biomass concentration using a microwave oven for drying. Chapter I presents the results with discussion of drying two samples in a microwave oven with settings of "defrost" and "cook" on the oven. The chapter also compares results of a growth curve constructed using dry weights calculated from conventional air oven drying and microwave drying. Each chapter of the thesis is self contained with references, nomenclature, tables and figures at the end of the chapter.

The second portion of the thesis concentrates on specific rate estimation and yield parameter estimation. Chapters II - IV discuss fitting smoothing spline functions to experimental data for the purpose of estimating rates and parameters. Chapter III introduces a new method, the response surface method, which makes use of spline functions in rate estimation. Chapter V concerns estimation of specific rates and introduces a procedure for selection of the region of exponential growth.

I would like to express my sincere thanks to my advisers and the graduate students in the chemical engineering department. A special thanks to Dr. L. Erickson and Dr. Mehmet Oner for their effort and enthusiasm.

I would also like to thank my fiance Rebecca Smith for her support, patience and love while I was writing.

This work was partially supported by the National Science Foundation.

M. Buono

TABLE OF CONTENTS

		Page
CHAPTER I.	RAPID MEASUREMENT OF CANDIDA UTILIS DRY WEIGHT WITH MICROWAVE DRYING	
	INTRODUCTION.	1.1
	MATERIALS AND METHODS	1.2
	RESULTS AND DISCUSSION.	1.5
	CONCLUSIONS	1.7
	REFERENCES.	1.7
	FIGURES	1.8
	TABLES.	1.9
CHAPTER II.	REDUCTION IN SCATTER PRESENT IN EXPERIMENTAL DATA AND ESTIMATION OF RATE QUANTITIES USING SMOOTHING SPLINE FUNCTIONS	
	INTRODUCTION.	2.1
	THEORY.	2.2
	MATERIALS AND METHODS	2.5
	RESULTS AND DISCUSSION.	2.5
	CONCLUSIONS	2.7
	REFERENCES.	2.8
	NOMENCLATURE.	2.10
	FIGURES	2.11

Page

CHAPTER III. COMPARISON OF TWO METHODS OF SELECTING SMOOTHING SPLINE FUNCTIONS FOR ESTIMATION OF SPECIFIC RATES IN FERMENTATIONS

INTRODUCTION 3.1

THEORY 3.2

MATERIALS AND METHODS. 3.7

RESULTS AND DISCUSSION 3.9

CONCLUSIONS. 3.13

REFERENCES 3.14

NOMENCLATURE 3.15

FIGURES. 3.16

TABLES 3.23

CHAPTER IV. CURVE FITTING EXPERIMENTAL BATCH/FED BATCH FERMENTATION DATA USING SMOOTHING SPLINE FUNCTIONS

INTRODUCTION 4.1

THEORY 4.1

MATERIALS AND METHODS. 4.4

RESULTS AND DISCUSSION 4.4

CONCLUSIONS. 4.9

REFERENCES 4.10

NOMENCLATURE 4.11

FIGURES. 4.12

TABLES 4.28

Page

CHAPTER V. ANALYSIS OF EXPONENTIAL GROWTH: ANALYSIS OF YIELD, TIME INTERVAL, AND MAXIMUM SPECIFIC GROWTH RATE.

INTRODUCTION 5.1

THEORY 5.1

MATERIALS AND METHODS. 5.7

RESULTS AND DISCUSSION 5.7

CONCLUSIONS. 5.9

REFERENCES 5.11

NOMENCLATURE 5.12

FIGURES. 5.14

TABLES 5.15

CHAPTER I

RAPID MEASUREMENT OF *CANDIDA UTILIS* DRY WEIGHT WITH MICROWAVE DRYING

INTRODUCTION

One significant parameter which describes the state of a bioreactor is the concentration of biomass in the reactor. There are many enumeration procedures available to determine cell concentration including turbidity, direct microscopic counting, dry weight analysis, counting with electronic particle counters, and dilution colony counting. Air oven drying of a fermentor sample is the technique most often used to indicate biomass concentration. However, this procedure requies 24 hr. or more before drying is complete. Therefore, for on line determination of biomass concentration in a bioreactor, which is desirable for effective control of the reactor, a quicker method for determining dry weights needs to be developed.

The purpose of this work is to investigate use of microwave drying to measure biomass concentration. Investigation into quick drying has been done for moisture determination of food and dairy products[1,2,3]. To the best of the authors' knowledge this is the first report on quick drying using a microwave oven to determine biomass concentration.

The power input to a sample in a microwave oven is greater than the energy input per unit time in an air oven, and therefore, less time is needed in a microwave oven. In order that the results in the chapter may be more clearly understood, a qualitative description of the different energy transport mechanisms involved in microwave drying and air drying is provided. In a conventional air oven the outer layer of the sample is heated by convection, while conduction heats the remainder of the sample. As a result, a temperature profile exists in a sample drying in an air oven at any one time. In a microwave oven, however, the radiant energy which is incident upon the sample penetrates through the sample and as a result the sample is uniformly

heated[1]. As a consequence, the sample experiences a higher evaporation rate than that which is experienced in an air oven. This is due to the higher average temperature which exists in the sample. In light of this, the results presented in this chapter are discussed in terms of the time necessary to measure cell concentration using a microwave oven. These results are compared with those obtained by conventional air oven drying.

MATERIALS AND METHODS

Power Rating of the Microwave

A Minutemaster microwave manufactured by Litton Microwave Cooking (Model # 403.002) was used in this study. The oven had only "cook" and "defrost" settings. On the "cook" setting, the oven is designed to use 1400 W of electrical power and provide 630 W of microwave power to heat the contents of the oven. On "defrost" the oven cycles on and off with 30 seconds duration in each position. The microwave power delivered to 75 ml of water was estimated by heating the water for two minute time periods and recording the change in temperature of the water.

Sample Exposure Time - Construction of Drying Curves

The length of time to expose a sample to microwave radiation in order to achieve drying of the sample which matched the drying obtained in an air oven for the same sample, was determined by constructing drying curves. A submerged culture of Candida utilis ATCC 1084 was sampled. Twenty ml samples were filtered through preweighed 1.2 μm type RA millipore filters (Millipore Corporation). One sample was covered with a second piece of filter paper to prevent sample splatter. The covered sample was then supported on an inverted 14 cm diameter pyrex watchglass, heater in the microwave for one

minute and then weighed.* The sample was heated another minute and then weighed once again. This procedure was continued until the loss in weight recorded was outside the range of accuracy of the mettler balance (about 0.01 mg.).

The second sample, which also was filtered through preweighed filter paper, was placed in an air oven to dry for 24 hr. at 105°C and then weighed. From these data drying curves were constructed.

Growth Curve

C. utilis was cultivated on YM broth (Difco Laboratories) in 250 ml flasks and incubated at 30°C in a Psychrotherm incubator manufactured by New Brunswick Scientific. Agitation was maintained at 200 rpm. A 10% (volume basis) inoculum was used. Each shake flask contained 150 ml of broth. Three 20 ml samples were taken from one flask each hour. One sample was used to obtain air oven dry weight; a second sample was used to obtain microwave oven dry weight; the third sample was a backup. Each shake flask was sampled twice.

Samples were dried for 15 min. in the microwave oven. The oven cooled for 45 min. between samples except for the sample at hour three. For the sample at hour 3.0, the oven had cooled for 1.75 hr. instead of for the usual 45 min. After 1.75 hr. had elapsed, drying was resumed. Samples taken after the sample at hour three were refrigerated until the next drying cycle.

*Some microwave ovens are not designed to operate with such a small microwave absorbing load.

RESULTS AND DISCUSSION

From the power measurement procedure described in the Materials and Methods section of this chapter, 150 W was found for the "defrost" setting and the "cook" setting provided heating at approximately 450 W. As described in the previous section of this chapter, drying curves were constructed from samples taken from a growing culture of C. utilis. A 10% error between air oven dry weights and microwave oven dry weights was deemed an acceptable result. This cutoff point was selected to insure that microwave oven dry weights could be easily correlated with air oven dry weights for standard curve construction.

The results are shown in Figures 1.1 - 1.4. All figures show sample weight as a function of exposure time. Figure 1.1 and 1.2 show the data collected for the small (14.42 mg. dry wt.) and large (147.74 mg. dry wt.) samples, respectively. With the oven set to "defrost", reduction in sample weight was best for the larger sample. After 7 min. of exposure the weight of the larger sample was within 7.7% of the air oven dry weight, while for the smaller sample 15 min. of exposure reduced the sample weight to within 24.2% of the air oven result. Therefore, small samples of about 14.4 mg., dried at the "defrost" power setting did not give reasonable results. However, for samples approaching 150 mg., dried at the "defrost" setting, acceptable results were obtained using microwave drying.

The results for the "cook" setting were better than those found for the "defrost" setting. Reduction in sample weight was best for the larger sample. The larger sample dry weight differed from the air oven value by 6.4% after only 3 min. of exposure. For the smaller sample the weight was reduced to within 9.2% of the dry weight value after 4 min. of exposure. Therefore, microwave drying gave acceptable results for both sizes when dried at the "cook" setting; however, for the "defrost" setting the results

were within 10% of the air oven result for only the larger sample.

To better appreciate the drying phenomena occurring in microwave drying, the quantity of water removed for each exposure was calculated. Figures 1.1 - 1.4 indicate that a larger quantity of water was removed at both power settings, for the larger sample size (147.74 mg. dry weight). The results are presented in Tables 1.1 and 1.2. With the oven set to "cook", 513.41 milligrams of water were removed during the first minute of exposure for the 147.74 mg. dry weight sample, while at the same power setting the smaller sample (14.42 mg dry weight) had 137.76 milligrams of water removed during the first minute. With the oven set to "defrost", similar results were observed, but the quantity of water removed was smaller, i.e., 453.18 and 93.64 mg. of water for the larger and smaller samples.

Several factors come into play as the size of the sample is increased. For large samples the rate of heat generation within the sample can exceed the rate of heat dissipation and result in sample charring[2].

Figure 1.5 shows growth curve results using both air oven dry weights and microwave oven dry weights. The greatest discrepancy observed between the two techniques was a difference of 0.16 g./L at hour 3.0. This most likely was due to a break in the drying and subsequent cooling cycle used in the oven. The microwave oven had less residual energy in the form of sensible heat present in the walls of the oven than was present in previous runs. Thus, there was a reduction in the quantity of water removed for this sample. Subsequent to this, however, the oven "heated up", and the quantity of water removed was again close to the air oven results. These results indicate that, the drying-cooling cycle used, is a parameter to which microwave drying is sensitive.

Many different microwave ovens are available. Some are designed to operate under the small sample conditions required in this work[2]; however, others should have a significant microwave absorbing load in them at all times to avoid damage to the oven and leakage of microwave energy from the oven. An appropriate absorbing load may be placed at a distant location within the oven.*

CONCLUSIONS

Microwave ovens can be used for quick drying of fermentor samples for the purpose of calculating biomass concentration. A reduction in measurement time from 24 hr. to 15 min. is observed with this technique. The necessary drying time is about 15 min. for the "cook" setting (approximately 630 W at full power). It is necessary to cover the sample with a second piece of filter paper to prevent loss of sample due to splatter.

Acceptable results were obtained for sample sizes with final dry weights ranging from 14.4 to 147.7 mg. at the "cook" setting. The large sample was reduced to within 2.5% of the air oven result after 15 exposures of one minute duration, while the smaller sample weight was within 7.1% of the dry weight result after 15 exposures.

REFERENCES

1. Barbano, D. M., and M. E. Della Valle, "Microwave Drying to Determine the Solids Content of Milk and Cottage Cheese," J. Food Prot., 47, 272 (1984).

2. Davis, A. B. and C. S. Lai, "Microwave Utilization in the Rapid Determination of Flour Moisture," Cereal Chem., 61, 1 (1984).

3. Minor, B. A., K. A. Sims, R. Bassette, and D. Y. C. Fung, "Comparison of Infradry and AOAC Methods for Moisture in Food Products," J. Food Prot., 47, 611 (1984).

4. Hayward, G. H. and D. H. Kropf, "Sample Position Effects on Moisture Analysis by a Microwave Oven Method," J. Food Prot., 43, 656 (1980).

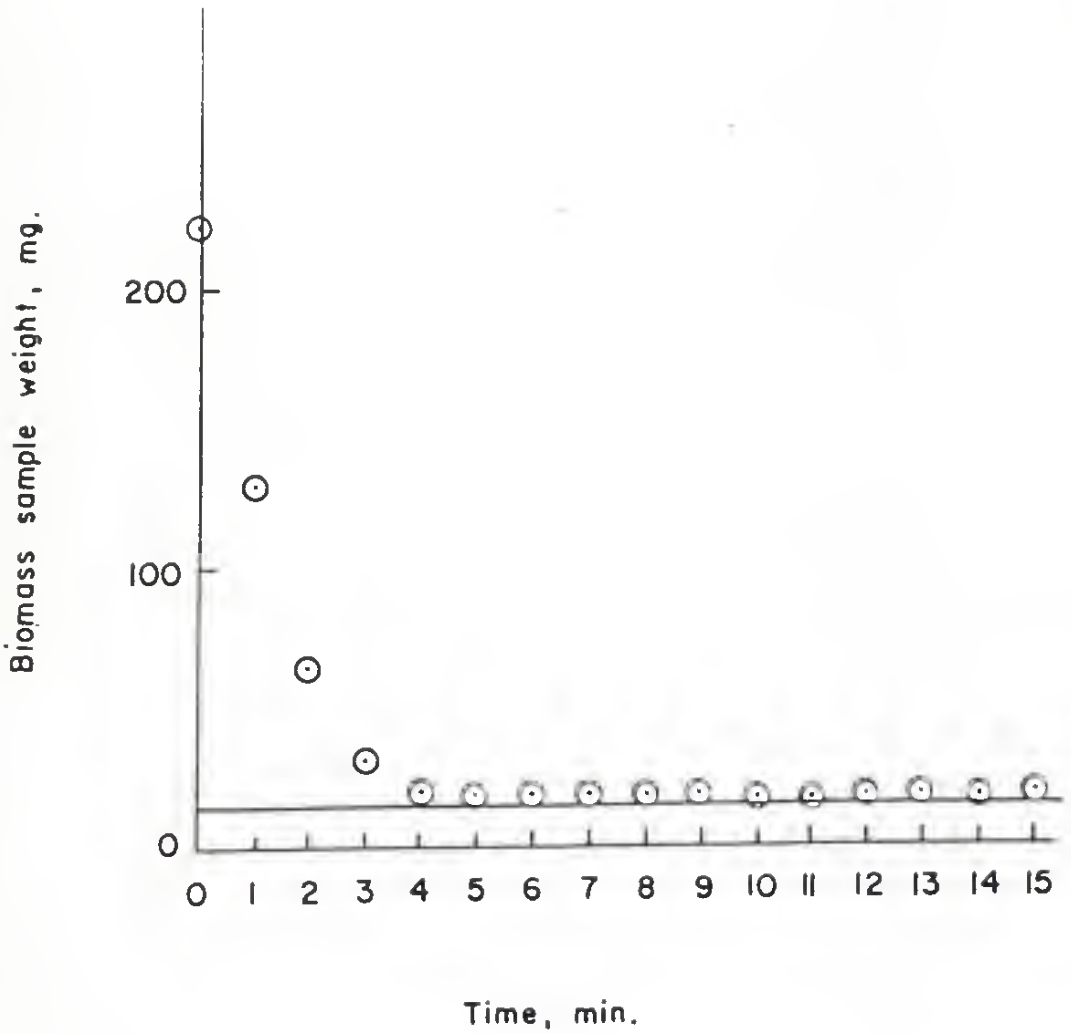


Figure 1.1 Drying curve for 14.42 milligram dry weight sample size with oven set to "defrost". The circles represent results obtained from microwave oven. The solid line is air oven final dry weight.

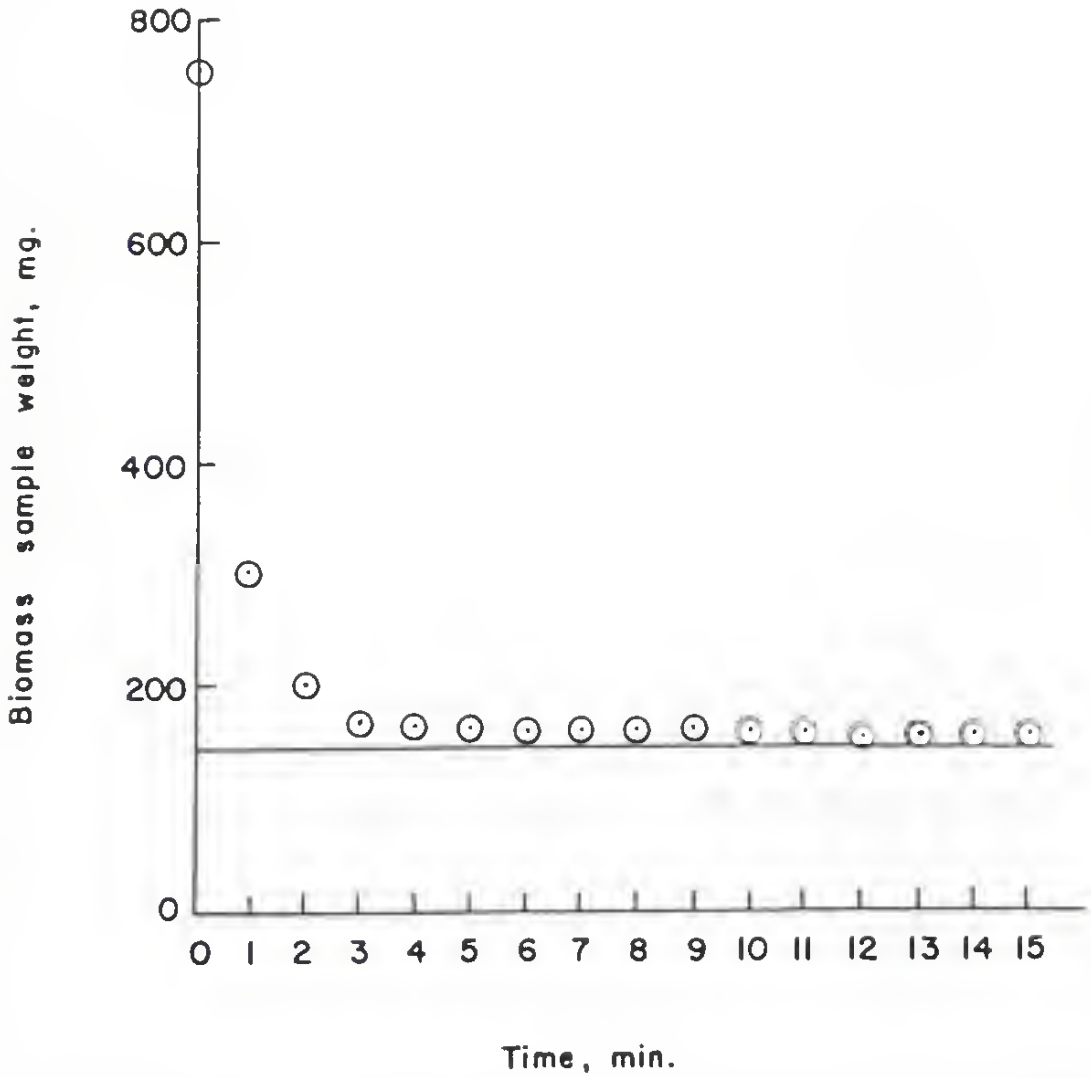


Figure 1.2 Drying curve for 147.74 milligram dry weight sample size with oven set to "defrost". The circles represent results obtained from microwave oven. The solid line is air oven final dry weight.

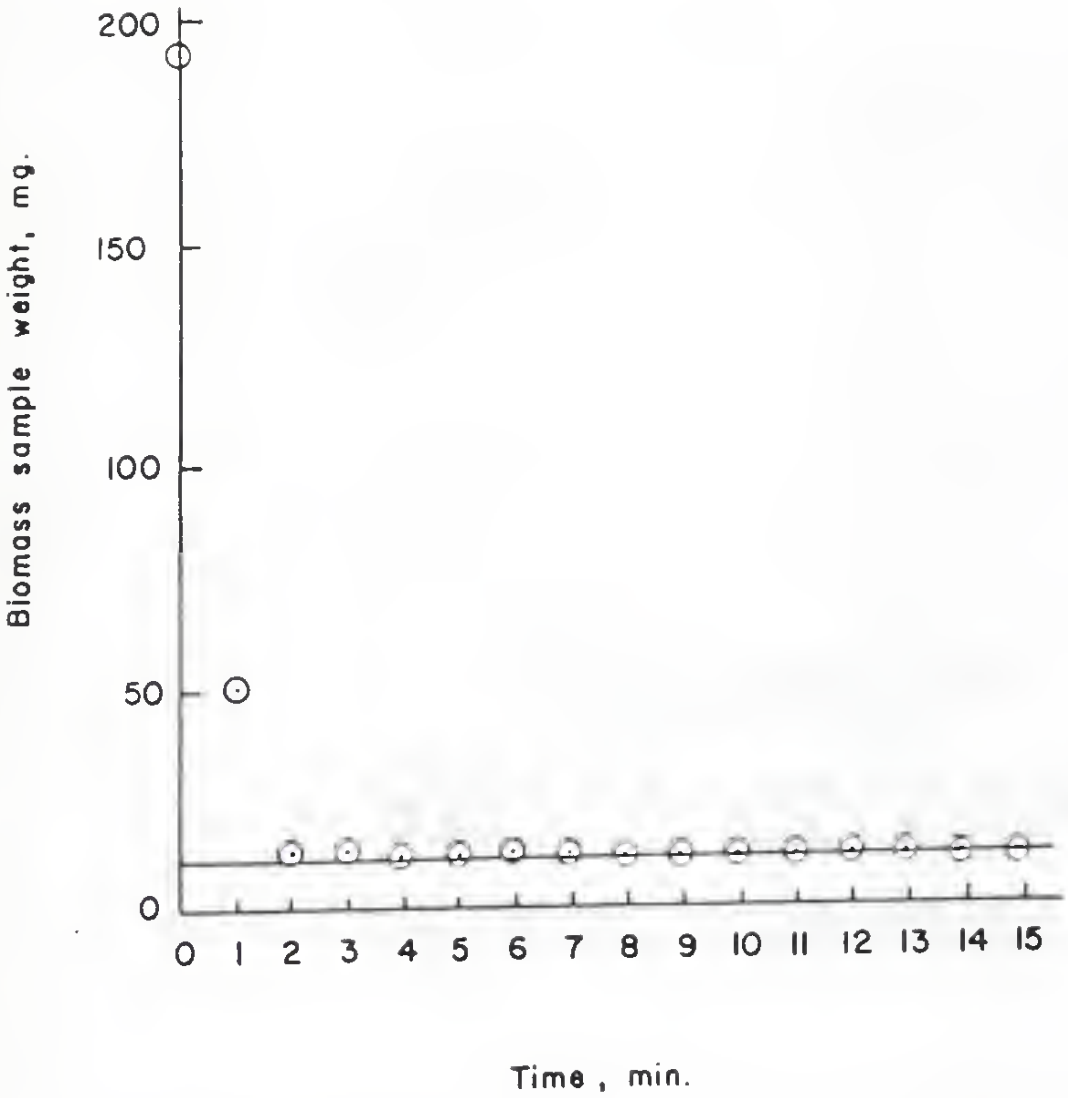


Figure 1.3 Drying curve for 14.42 milligram dry weight sample size with oven set to "cook". The circles represent results obtained from microwave oven. The solid line is air oven final dry weight.

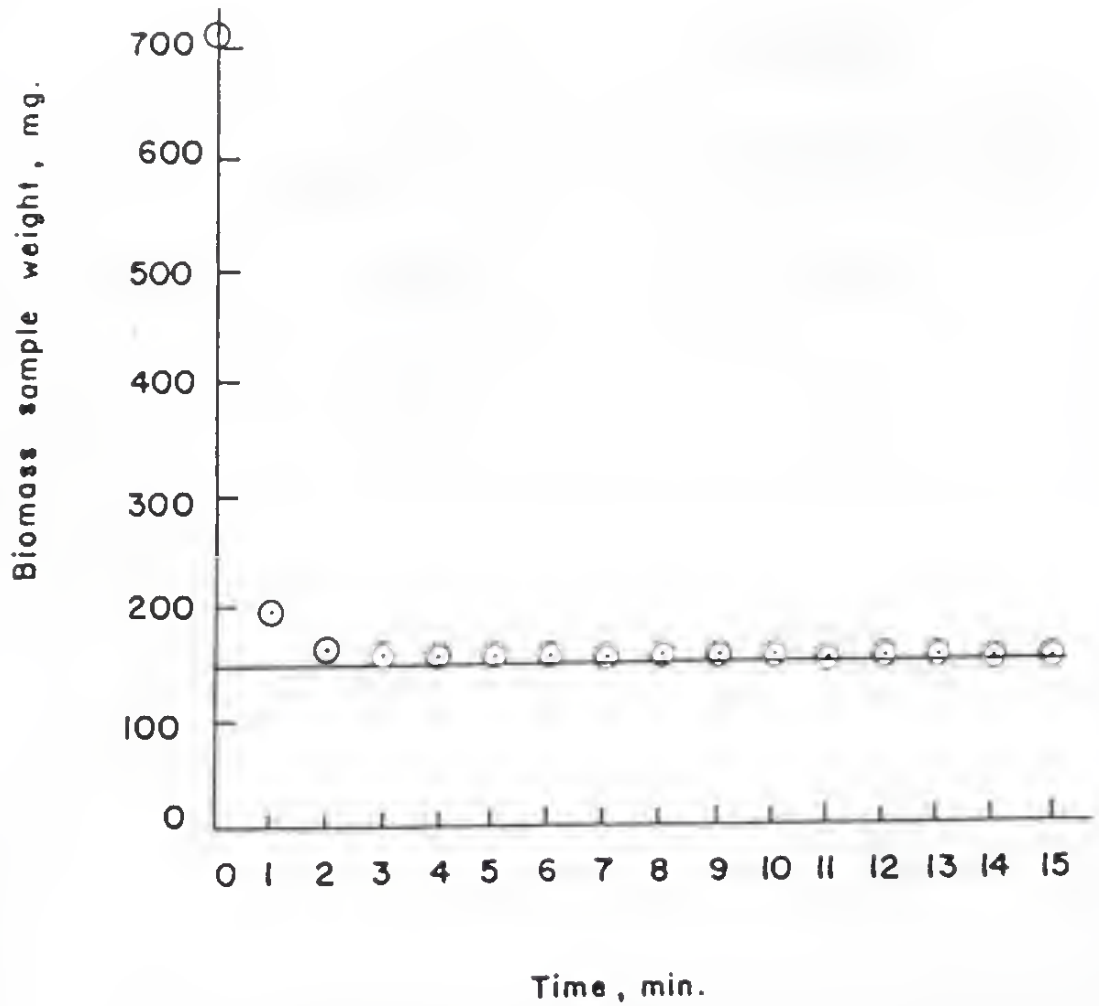


Figure 1.4 Drying curves for 147.74 milligram dry weight sample size with oven set to "cook". The circles represent results obtained from microwave oven. The solid line is air oven final dry weight.

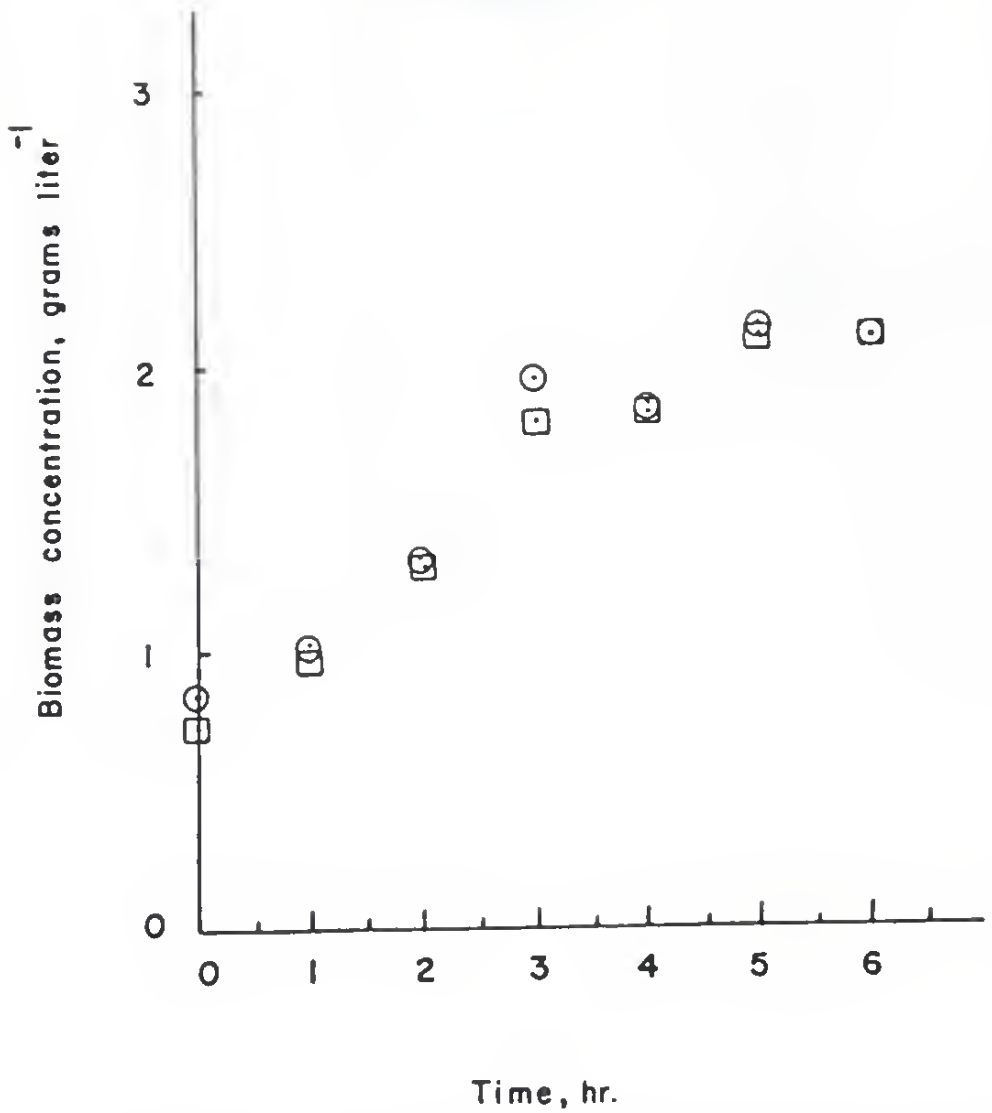


Figure 1.5 Growth curve for Candida utilis grown on glucose. The circles represent microwave dry weights obtained for the "cook" setting. The squares represent air oven dry weights.

Table 1.1. Water Removed per Exposure for the First Nine Exposures of One Minute Duration for "Defrost" Power Setting.

Exposure Time Interval, min.	Water Removed, mg.	
	Small Sample	Large Sample
0.0-1.0	93.64	453.18
1.0-2.0	64.79	101.84
2.0-3.0	34.46	30.72
3.0-4.0	10.08	6.49
4.0-5.0	0.96	1.89
5.0-6.0	0.00	0.33
6.0-7.0	0.00	1.06
7.0-8.0	0.43	0.93
8.0-9.0	0.00	0.00

Table 1.2. Water Removed per Exposure for the First Fourteen Exposures of One Minute Duration for the "Cook" Power Setting.

Exposure Time Interval, min.	Water Removed, mg. Small Sample Size, 14.42 mg. dry wt.	Large Sample Size, 147.74 mg. dry wt.
0.0-1.0	137.76	513.41
1.0-2.0	37.52	33.22
2.0-3.0	1.45	5.08
3.0-4.0	2.80	1.61
4.0-5.0	0.21	0.56
5.0-6.0	0.00	0.90
6.0-7.0	0.05	0.49
7.0-8.0	0.00	0.78
8.0-9.0	0.27	0.06
9.0-10.	0.00	0.07
10.-11.	0.23	0.17
11.-12.	0.00	0.19
12.-13.	0.21	0.32
13.-14.	0.00	0.00

CHAPTER II

REDUCTION IN SCATTER PRESENT IN EXPERIMENTAL DATA AND ESTIMATION OF RATE QUANTITIES USING SMOOTHING SPLINE FUNCTIONS

INTRODUCTION

In the absence of product formation, estimation of maintenance and yield coefficients and determination of the state of a bioreactor depend upon values of the specific growth rate, μ, and the specific rate of substrate utilization, Q_s. For precise estimation of these variables and for accurate determination of the state of a bioreactor, precise estimates of these rate quantities need to be known. Since there is a large amount of water present in fermentation systems, it is difficult to construct a hydrogen balance and an oxygen balance. However, measurement techniques are available, which permit a carbon balance and an available electron balance to be made on these systems. Best estimates of the specific growth rate and the specific rate of substrate utilization obtained may be based on how well these balances are satisfied. The rate quantities associated with μ and Q_s are calculated from the time derivatives of the biomass and substrate concentration curves, respectively. The first goal of this work was to address how these rate terms can be consistently calculated. Once this has been determined, refinement of rate estimation can proceed by examining how well the carbon balance and the available electron balance are satisfied for the μ_i and Q_{si} calculated.

All measurements made on a chemical reactor possess error. As a result, there is a certain degree of scatter associated with the data obtained from these measurements. For the purpose of rate estimation, derivatives are sometimes calculated from tangents drawn to curves fitted to the data by hand smoothing. However, fitting the curve by hand is not desirable statistically since the shape of the curve which is drawn is dependent on the draftperson who fits the curve. As a consequence, it is difficult to consistently fit a smooth curve through the data. Polynomials are sometimes used to fit data; however, a polynomial is influenced in one region by its shape in another

region. Physical data do not follow this type of pattern. As stated by Wold,[1] physical data are more often disjointed in nature, i.e., the behavior of the data in one region may not be related to the behavior of the data in another region. Spline functions which are piecewise polynomials provide a solution to the problem of curve fitting experimental data.

The second goal of this work was to demonstrate reduction in the scatter present in experimental data by fitting curves to data with smoothing spline functions. The reduction in scatter is displayed by comparing the rate terms calculated from the spline functions with the derivative quantities calculated from hand fit curves. Upon differentiation of experimental data error present in the data becomes enhanced, as a result, a comparison of the rate terms was deemed an acceptable method to test spline functions for their smoothing ability in curve fitting experimental data.

THEORY

A spline function is defined as a piecewise polynomial of the n^{th} order which is a continuous function with (n-1) continuous derivatives. A spline function represents experimental data quite well because of its piecewise form; the function is fitted through "knots", (a "knot" ties a spline function at the first and last values of the independent variable for which the spline function applies) and for this study each data point was selected as a "knot". Since, a spline function works independently within each region, i.e., between two "knots", gives a more accurate representation of the underlying function than does an ordinary n^{th} order polynomial. As developed by Reinsch[2], a cubic spline takes the form,

$$g(t) = a_i + b_i (t-t_i) + c_i (t-t_i)^2 + d_i (t-t_i)^3 \qquad (2.1)$$

for $t_i < t < t_{i+1}$

A smooth spline function is a spline function, g(t), which minimizes the following equation:

$$P = \sum_{i=1}^{N} (((g(t) - y_i)/dy_i)^2 + \lambda \int_{t_o}^{t_N} (g''(t))^2 dt) \qquad (2.2)$$

In the above equation, the spline function fit to the data appears in two terms: in the form of its zero and second order derivatives, respectively. The first term insures fidelity of the spline fit to the data, and the second term determines the smoothness of the fit, or the magnitude of the curve's curvature. The second derivative of a function provides an estimate of the degree of smoothness of the function and, as a result, in Eq. (2.2), as λ is increased the intergral term becomes more important. Therefore, the spline function will be smoother. In the limit $\lambda \longrightarrow \infty$, the least squares straight line through the data results. Consequently, as λ is decreased the integral of the square of the second derivative becomes less important in Eq. (2.2), the smoothness of the spline function decreases. In the limit $\lambda \longrightarrow 0$ the smoothing spline function reduces to an interpolating spline function, and the curve fit to the data passes through each point.

As described above, λ controls the "degree of smoothness" of the curve fit to the data. The smooth function, g(t), to be constructed shall minimize $\int (g''(t))^2 dt$ under the constraint

$$S \geq \sum_{i=1}^{N} (y_i - g(t))^2/(dy_i)^2 \qquad (2.3)$$

Wabha and co-workers[3] stat that if g(t) is the minimizing function under the constraint (2.3), then there exists $\lambda=\lambda(S)$ for which g(t) is the spline function which minimizes Eq. (2.2). For the discussion that follows below, reference to the "degree of smoothness" of a smooth spline function fit to data is reference to S.

MATERIALS AND METHODS

The IMSL subroutine ICSSCU was used to generate smoothing spline functions which satisfied Eq. (2.2). The subroutine requires the user to specify a value for S, the "degree of smoothness", and also to give an estimate of the relative weight of each data point, $1/dy_i$. Selection of the optimal value for S is discussed in Chapter III. In this study, all data points were equally weighted; computer simulations indicate that equal weighting of the data points provide the most consistent estimates of biomass concentration, substrate concentration, the specific rate of microbial growth, and the specific rate of substrate utilization, from the viewpoint of satisfying the carbon balance and the available electron balance. An estimate for the standard deviation of the data, σ, provided the value for the weight which was assigned to each point. σ was calculated from a curve which was hand fit to the data to be smoothed. The differences between the experimental values and those predicted by the hand fit curve were used to compute the standard deviation.

The data examined and the results evaluated were those of Solomon and coworkers[4]. The results obtained herein were compared with those of Solomon.

RESULTS AND DISCUSSION

The biomass and substrate concentration curves fit with smoothing spline functions for different smoothing parameter values are plotted along with the experimental points in Figures 2.1 - 2.4. The experimental data were collected from a bioreactor operated in batch mode up to 6.0 hours; at hour 6.5 feeding commenced. Figures 2.1 and 2.2 show that maximum function curvature resulted with S set equal to zero. For this case, the data were interpolated.

The IMSL User's Manual suggests that the smoothness parameter, S, for the smooth spline function used lie within the confidence interval of the left hand side of Eq. (2.2). This is given by the following expression,

$$N - (2N)^{1/2} < S < N + (2N)^{1/2} \qquad (2.4)$$

For the data set of Solomon[4], N was equal to twenty-eight. However, due to the large scatter present in the data at the point where fed batch operation began, smooth curves were fit to the data with S set equal to 100.

The results, plotted in Figures 2.3 and 2.4 show that spline estimates of biomass and substrate concentration through the interval where fermentor operation was switched from batch to fed batch trace smooth functions. Further, the spline functions fit through all the data show little fluctuation. As a consequence, the time derivatives evaluated from the spline functions also possessed little scatter. This is shown in Figures 2.5 and 2.6. These plots display biomass concentration time derivatives and substrate concentration time derivtives calculated from curves hand fit to the experimental data and, the time derivatives calculated from the smoothing spline curve fit to the data with S set equal to 100. As discussed by Solomon and co-workers[4], derivatives were calculated from the hand smooth curves using finite differences. The results obtained from this procedure show a significant amount of scatter when compared with the results obtained with spline functions for both the batch and fed batch regimes. Thus, we concluded that smoothing spline functions provided a means for reducing the variability present in the experimental data and in the rate quantities calculated from these data.

To further examine the smoothing spline function, values for the specific growth rate, the specific rate of substrate utilization and the biomass

energetic yield were calculated. These were determined from the time derivatives of the spline function fit to the biomass concentration and substrate concentration data, respectively. Experimental values for biomass and substrate concentrations were used in the calculations, according to the expressions

$$\mu = 1/X_{Exp}\ dg/dt + D \tag{2.5}$$

$$Q_s = 1/X_{Exp}\ df/dt + D(S_o - S)/X_{Exp} \tag{2.6}$$

where g is the spline function fit to the biomass concentration data and f is the spline function fit to the substrate concentration data. Point values of the biomass energetic yield were calculated, viz.,

$$\eta = (\sigma_b\ \gamma_b\ \mu)\ /\ (Q_s\ \sigma_s\ \gamma_s) \tag{2.7}$$

The results obtained for both the batch and fed batch mode of fermentor operation are shown in Figure 2.7. Figures 2.5, 2.6 and 2.7 show that indeed smooth estimates of derivatives based on quantities from experimental data possessing scatter were possible with spline functions.

CONCLUSIONS

Spline functions have been used successfully to smooth experimental data; where a significant amount of scatter was present in the biomass data and the spline function fit a smooth curve to the data.

The smoothing spline function fit to the data varied significantly with the extent of smoothing selected. For the smallest value of the smoothing parameter, i.e., S=0, the smoothing spline function reduced to an interpolating spline function. However, for a larger value, i.e., S=100, virtually all scatter in the data was eliminated. As a result, upon differentiation, smooth estimates of rate quanitites were determined.

Complete reproducibility of rate estimation is possible with smoothing spline functions. Therefore, spline functions are an excellent technique for use in calculation of the specific rate of microbial growth and the specific rate of substrate utilization.

REFERENCES

1. Wold, S., "Spline Functions in Data Analysis," Technometrics, 16, (1974).

2. Reinch, C. H., "Smoothing By Spline Functions," Numerische Mathematik, 10, 177 (1967).

3. Wahba, G., and S. Wold, "A Completely Automatic French Curve: Fitting Spline Functions by Cross Validation," Communications in Statistics, 4, 1 (1975).

4. Solomon, D. O., L. E. Erickson, and S. S. Yang, "Utilization of Statistics and Experimental Design in Data Collection and Analysis," Biotechnol. Bioeng., 25, 2683 (1983).

NOMENCLATURE

a_i, b_i, c_i, d_i = spline function coefficients

D = dilution rate

dy_i = weight of the i^{th} observation

$g(t)$ = spline function fitted through data point t

$g''(t)$ = second derivative of the smoothing spline function

N = number of experimental observations

Q_s = the specific rate of substrate utilization

S = smoothing parameter

s = substrate concentration in the fermentor

s_0 = substrate concentration in the feed

X_{Exp} = experimentally measured biomass concentration

y_i = experimental observations

γ_b = reductance degree of the biomass

γ_s = reductance degree of the substrate

η = the biomass energetic yield coefficient

λ = smoothing parameter

μ = the specific rate of microbial growth

σ = standard deviation of the experimental observations

σ_b = weight fraction carbon in biomass

σ_s = weight fraction carbon in substrate

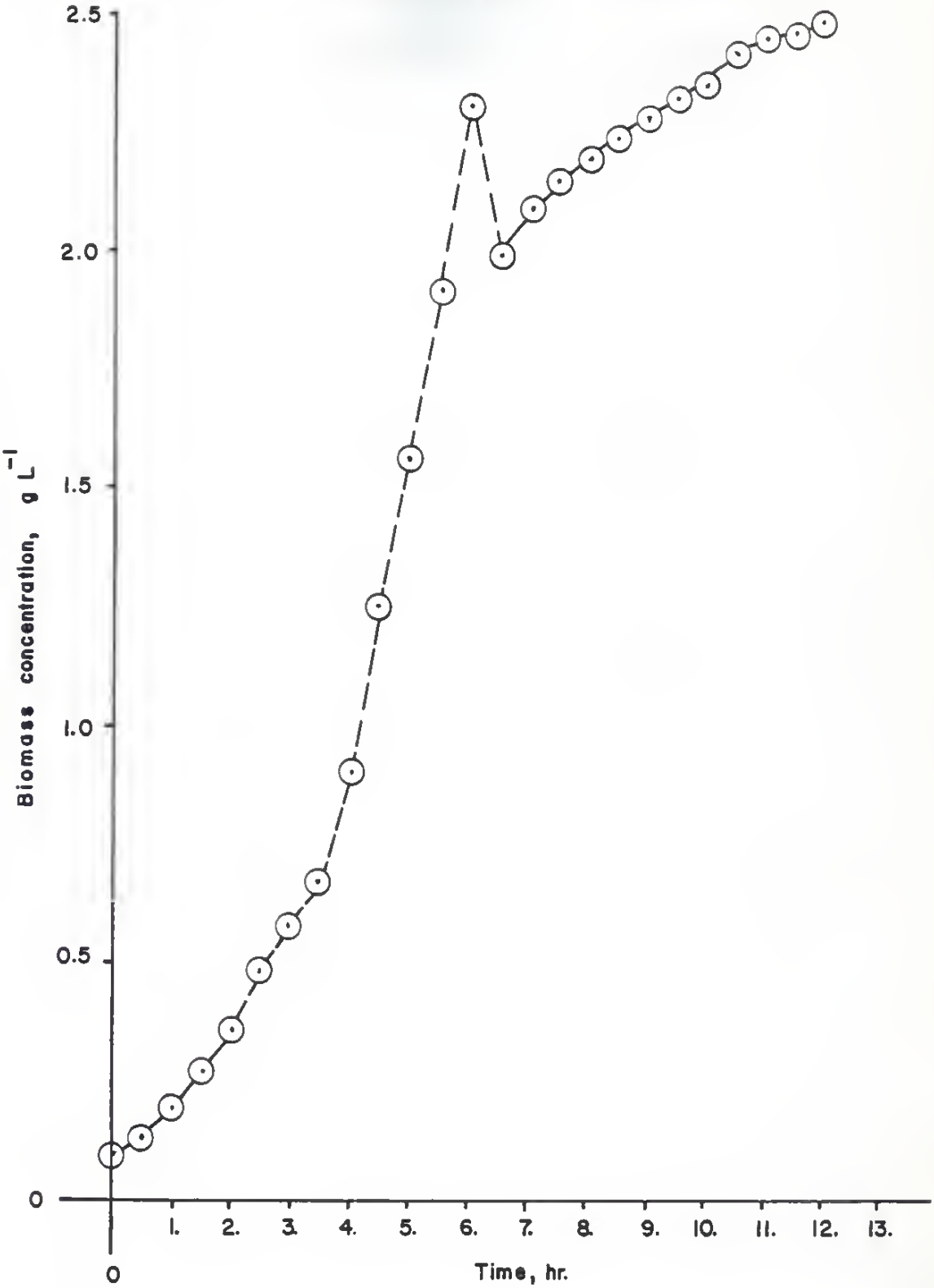


Figure 2.1 The smoothing spline function fitted to biomass concentration versus time data with S=0. The dashed line represents the spline function, and the circles represent the experimental data.

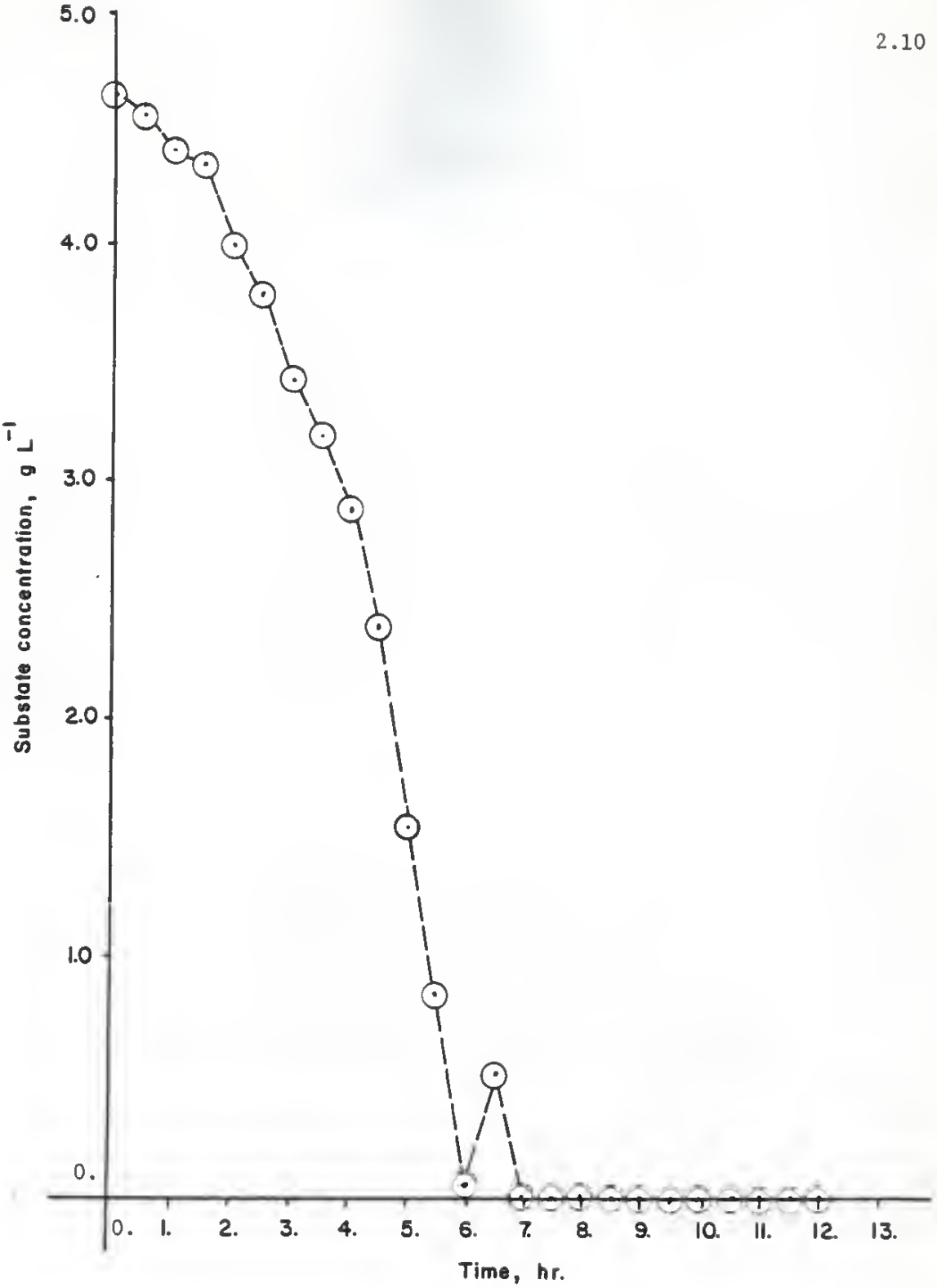


Figure 2.2 The smoothing spline function fitted to substrate concentration versus time data with S=0. Dashed line represents the spline function. Circles represent the experimental data.

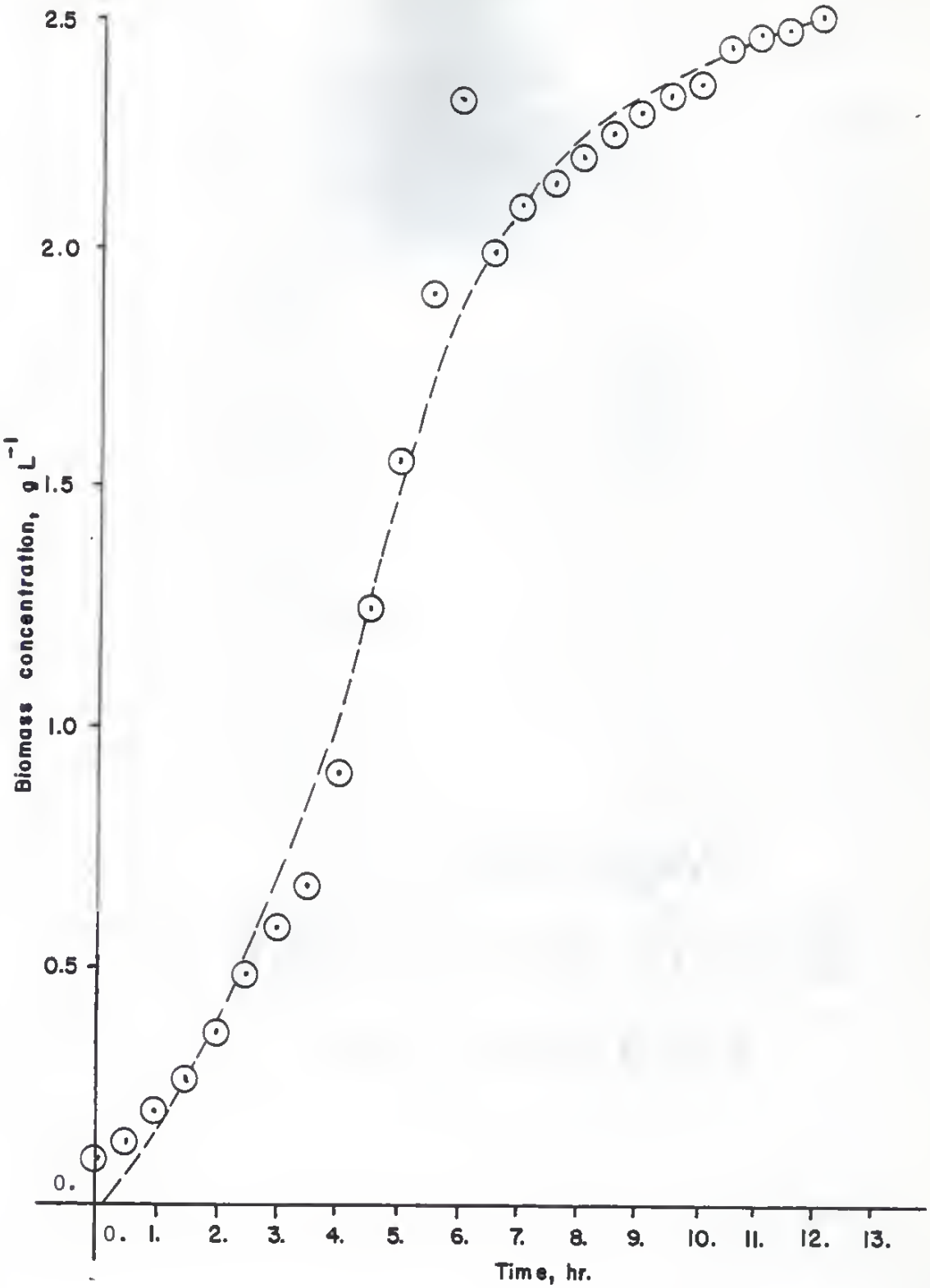


Figure 2.3 The smoothing spline function fitted to biomass concentration versus time data with S=100. The dashed line represents the spline function. The circles represent the experimental data.

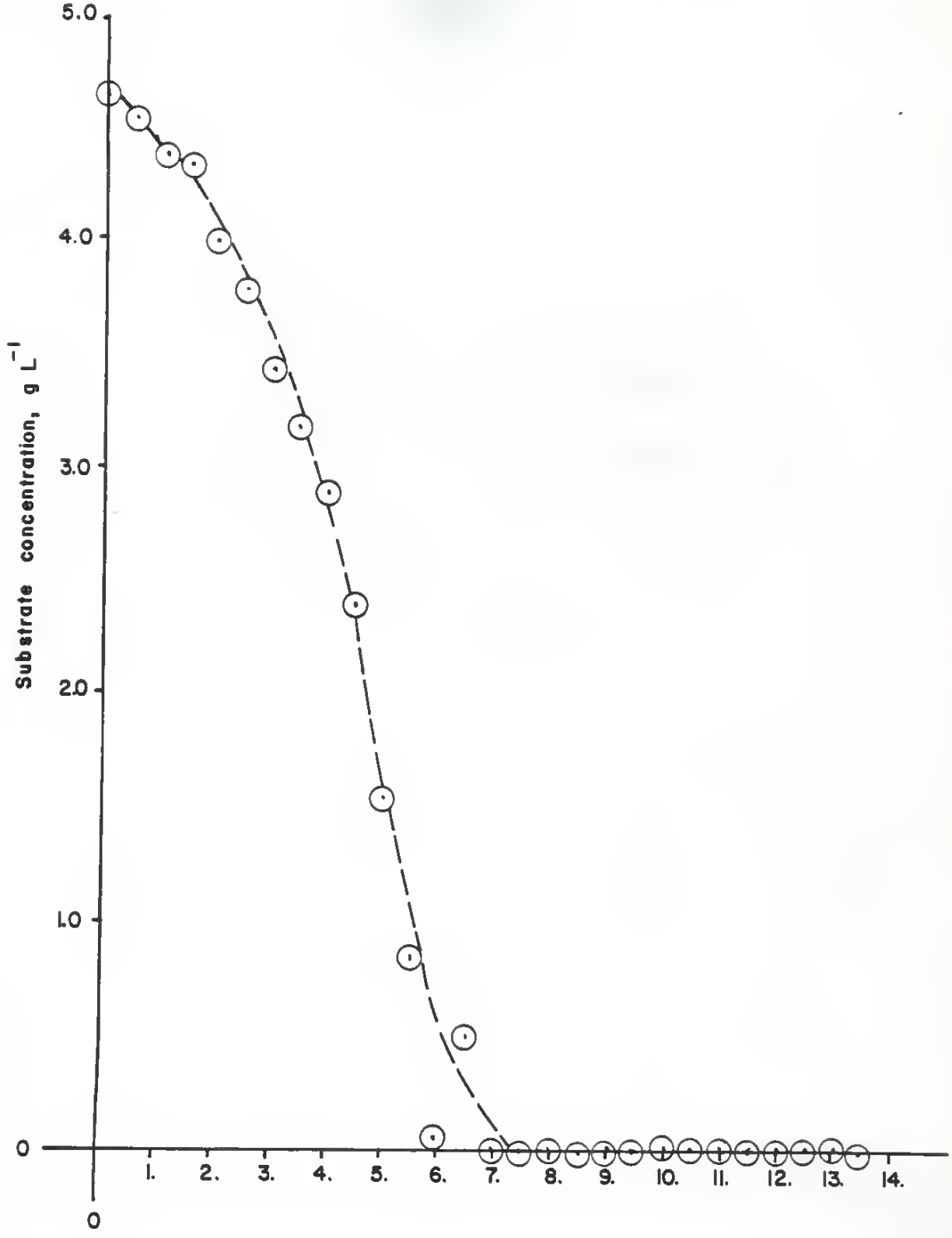


Figure 2.4 The smoothing spline function fitted to substrate concentration versus time data with S=100. The dashed line represents the spline function. Circles represent the experimental data.

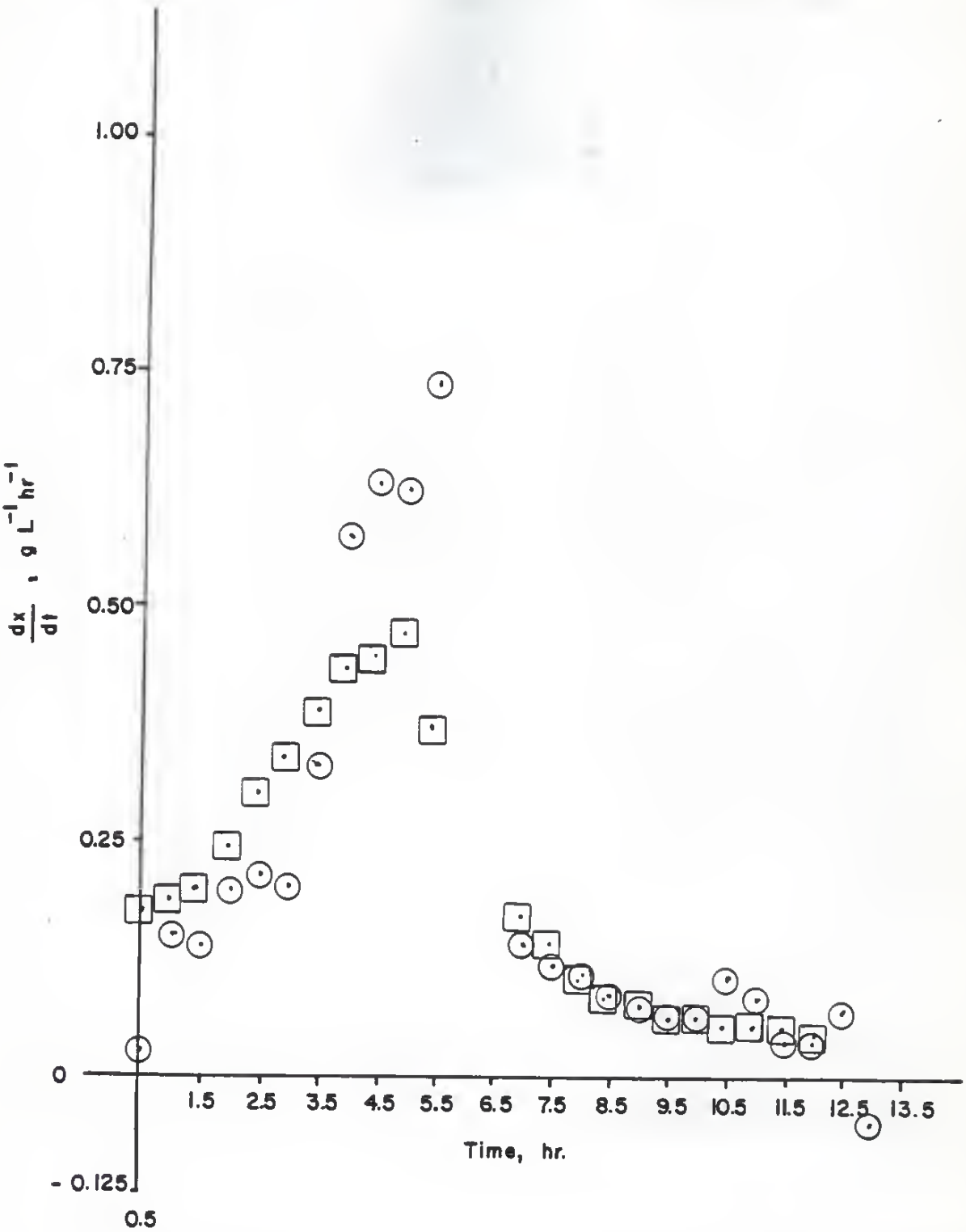


Figure 2.5 The time of change of biomass concentration plotted versus time. Boxes represent derivatives calculated from the smoothing spline function with S=100. Circles represent derivatives calculated from finite differences of a curve hand fitted to the experimental data.

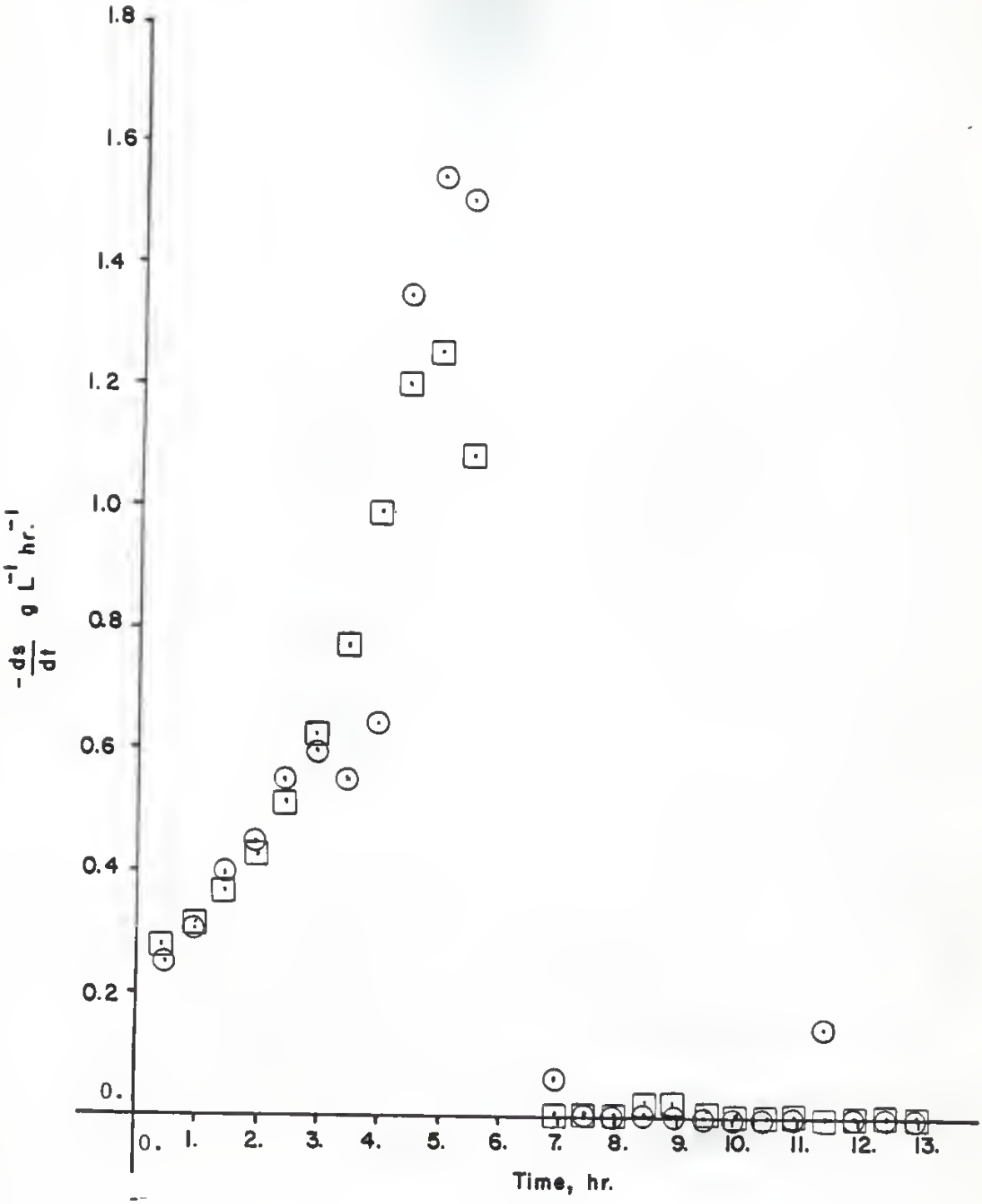


Figure 2.6 The time rate of change of substrate concentration plotted versus time. Boxes represent derivatives calculated from the smoothing spline function with S=100. Circles represent derivatives calculated from finite differences of a curve hand fitted to the experimental data.

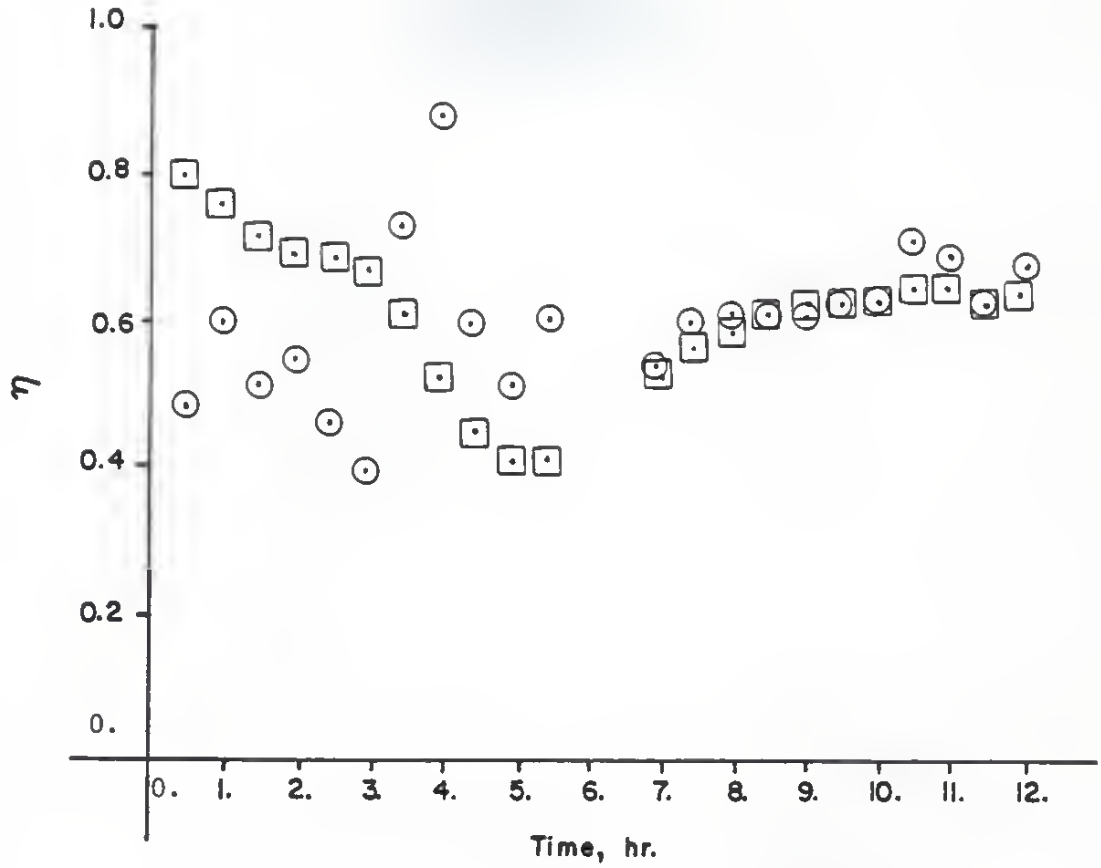


Figure 2.7 Estimates of the biomass energetic yield as a function of time. Boxes represent the estimates obtained using a spline function to calculate rate terms. Circles represent the estimates obtained from the μ_i and Q_{si} calculated using finite differences to determine rate quantities.

CHAPTER III

COMPARISON OF TWO METHODS OF SELECTING SMOOTHING SPLINE FUNCTIONS FOR ESTIMATION OF SPECIFIC RATES IN FERMENTATIONS

INTRODUCTION

Chapter II discusses the use of smoothing spline functions to fit curves to experimental fermentation data. As discussed, the smoothness parameter of the smoothing spline function is best viewed as an independent variable. Chapter II shows that the value of the smoothness parameter, S, greatly affects the curve fitted to the data. The primary objective of this work is to develop a procedure by which S is selected. A response surface method which is based on satisfying the carbon balance and the available electron balance is presented.

The mathematics of smoothing spline functions has been extensively developed by Reinsch[1], Wold[2] and Wahba and Wold.[3] The smoothness parameter, S, can be selected using cross validation as discussed by Silverman.[4] However, cross validation has been shown to require extensive computation. Further, selection of S is based solely on the experimental data to which the spline function is fit. The response surface procedure presented here is an effort to consider the consistency of the results, using the carbon and available electron balances, when selecting the spline function which best fits the data. All the measurements are considered for the response surface method while the cross validation method considers only the data for the curve that is being fitted.

There is currently considerable interest in the automation of the data collection and analysis processes associated with fermentations. This work provides another method of smoothing experimental data and evaluating the quality of the spline function fit of the data.

THEORY

Factorial Experimental Design and Response Surface Method

Experimental designs and response surface methods have been described by Box and Hunter[5], Box and Wilson[6] and Box.[7]

The experimental design used in this study to develop the response surface was the central composite design as described in Myers[8] with transformation of the values for the independent variables to ones, zeroes, and alphas where zero represents the origin and +1 represents the upper bound and -1 the lower bound for the independent variables. For two independent variables the design matrix consists of the following transformed sets of values of the independent variables:

$$(-1,-1),\ (1,-1),\ (-1,1),\ (1,1) \tag{3.1}$$

A central composite design permits the second order response surface to be fitted using a first order factorial design augmented by the following additional values of the two independent variables;

$$(0,0),\ (0,\alpha_2),\ (0,-\alpha_2),\ (\alpha_1,0),\ (-\alpha_1,0) \tag{3.2}$$

In the application of the response surface method to smoothing the biomass data and substrate data, S_1 and S_2 are the smoothing parameters for fitting the biomass and substrate data, respectively, with spline functions. The recommended range of values for S_1 and S_2 is given by the confidence interval

$$N-(2N)^{\frac{1}{2}} < S < N+(2N)^{\frac{1}{2}} \tag{3.3}$$

where N is the number of data points. A window of values of S_1 and S_2 can be defined based on this interval. Values of α_1 and α_2 were selected such that $\alpha_1 = \alpha_2 = 1.25$.

The response function is a measure of the sum of squares of the deviations of the carbon and available electron balances from one, that is,

$$F = \frac{1}{N} \sum_{i=1}^{N} [(J_{ci} - 1)^2 + (J_{ai} - 1)^2] \tag{3.4}$$

where

$$J_{ci} = \frac{\frac{\mu_i \sigma_b}{12} + Q_{ci}}{\frac{\sigma_s Q_{si}}{12}} \tag{3.5}$$

and

$$J_{ai} = \frac{\frac{\mu_i \sigma_b \gamma_b}{12} + 4\, Q_{oi}}{\frac{\sigma_s \gamma_s Q_{si}}{12}} \tag{3.6}$$

In the above equations, the specific growth rate, μ, is determined using the experimentally measured biomass concentration, X, in the expression (Solomon *et al.*[9])

$$\mu = D + \frac{1}{X} \frac{dX}{dt} \tag{3.7}$$

which may be used for both batch and fed batch cultures. The spline function representation of the biomass data is used to estimate dX/dt. The specific rate of substrate consumption, Q_s, is estimated using the expression (Solomon *et al.*[9])

$$Q_s = \frac{D(S_F - S)}{X} - \frac{1}{X} \frac{dS}{dt} \tag{3.8}$$

the experimentally measured substrate concentration, S, and the spline function representation of dS/dt.

The spline function fitting of each set of experimental data was by the response surface method.

This method is designated as Method I in this work.

Cross Validation (Method II)

The cross validation method is a second method for selection of optimal values for the smoothness parameters, S_1 and S_2, for biomass and substrate concentration data. One data point at a time is deleted from the data set and a spline function is fitted to the remaining data of the set for a given value of S. The error at the deleted point i

$$[y_i - g_s\ (t_i)]$$

is combined with the error associated with each of the other N-1 spline functions at each of the other N-1 deleted points to obtain[5]

$$XVSC = \frac{1}{N} \sum_{i=1}^{N} [y_i - g_s\ (t_i)]^2 \qquad (3.12)$$

where $g_s\ (t_i)$ is the spline function for value S of the smoothing parameter and point i deleted from the data set. The optimal value of S is selected as the value which minimizes Equation (3.12). With this method each curve is fitted independently.

Estimation of Growth Yield and Maintenance Coefficient

Once the derivative quantities have been estimated using spline functions, the specific growth rate, μ, and the specific substrate consumption rate, Q_s, may be estimated using Equations (3.7) and (3.8). The growth yield is

$$\eta = \frac{\sigma_b \gamma_b \mu}{\sigma_s \gamma_s Q_s} \qquad (3.13)$$

Pirt's model[10] may be written in the form (Solomon et al.[9])

$$\frac{1}{\eta} = \frac{1}{\eta_{max}} + \frac{m_e}{\mu} \tag{3.14}$$

and in the form

$$\frac{\mu}{\eta} = \frac{\mu}{\eta_{max}} + m_e \tag{3.15a}$$

The carbon balance, Equation (3.5), and the available electron balance, Eq. (3.6), may be incorporated into Eq. (3.15a) to obtain (Solomon *et al.*[9])

$$\frac{\mu J_a}{\eta} = \frac{\mu}{\eta_{max}} + m_e \tag{3.15b}$$

and

$$\frac{\mu J_c}{\eta} = \frac{\mu}{\eta_{max}} + m_e \tag{3.15c}$$

Point estimates for η_{max} and m_e are determined from Eqs. (3.15a), (3.15b) and (3.15c) by a least squares fit of μ/η on μ, $(\mu/\eta)(J_a)$ on μ and $(\mu/\eta)(J_c)$ on μ, respectively. In theory, Eqs. (3.15a), (3.15b) and (3.15c) are identical; however, each of the above three equations yields different estimates for $1/\eta_{max}$ and m_e because the carbon balance and the available electron balance are not identically equal to one. There has been considerable work recently (Yang *et al.*[11]; Solomon *et al.*[9]) using the covariate adjustment procedure of data analysis to determine best estimates for η_{max} and m_e using Eqs. (3.15a), (3.15b) and (3.15c).

MATERIALS AND METHODS

The experimental data analyzed were the batch and fed batch data of 5-3-82 of Solomon et al.[9] The data set contained twenty-eight observations, i.e., N = 28.

Method I

The response surface method is available as part of SAS[12]. By conducting a two dimensional optimization of Eq. (3.11) the optimal smoothness parameter for generation of the biomass curve, S_1, and the optimal smoothness parameter for generation of the substrate curve, S_2, were calculated. Several methods were investigated for this purpose: differential calculus, the method of sucessive variation of parameters, a quasi Newton method and an optimization algorithm available in SAS. Differential calculus and the method of sucessive variation of parameters were used in the study. The technique which was used most extensively for function minimization was differential calculus. In the case of occurrence of a saddle point, values for S_1 and S_2 determined from differential calculus provided initial guesses for subsequent optimization by a direct search.

For the data set studied, starting with the interval established for S by Eq. (3.3), by trial and error window; of values for S_1 and S_2 were centered about 60. For each parameter, the upper bound was set to 100 and the lower bound was set to 20. The axial distance was set to 50. The composite design determined the points used to generate values for Eq. (3.4); that is, S_1 and S_2 were set to (20,20), (100,20), (20,100), (100,100), (60,60), (60,110), (60,10), (110,60) and (10,60) according to the design. The IMSL subroutine ICSSCU was used to fit cubic splines to biomass and substrate concentration data. From the spline function fit, values of the derivatives

were calculated. The specific growth rate and the specific rate of substrate consumption were calculated using the experimental concentrations, the derivatives from the spline functions, and Eq. (3.7) and (3.8).

Using the remainder of the data, the specific rate of oxygen uptake and the specific rate of carbon dioxide evolution were evaluated. Data consistency was determined by evaluation of Eqs. (3.5) and (3.6). Equation (3.4) was solved to give the response to the values of S_1 and S_2. This procedure determined one point of the response surface. The parameters in Eq. (3.11) were estimated using the nine responses and SAS.

Method II

The IMSL subroutine ICSSCV was used to generate smoothing spline functions using cross validation. Values of the specific growth rate and the specific substrate consumption rate were estimated using the experimental concentrations and derivatives obtained from the spline functions.

Two sets of estimates for the specific growth rate, the specific rate of substrate consumption, the carbon balance, the available electron balance, and the growth yield η were calculated from Eqs. (3.5), (3.6), (3.7), (3.8) and (3.13) using the two methods.

The function PROC GLM of SAS was used to fit these results to Eq. (3.15). The slope was set equal to $1/\eta_{max}$, and the intercept was set equal to m_e. 95% confidence intervals of the point estimates were determined from the standard error of the estimate calculated by SAS. The covariate adjustment method was also used in this work to estimate the parameters using all the data simultaneously.

RESULTS AND DISCUSSION

The response function, Eq. (3.11), which relates the closure of the carbon balance and the available electron balance to the smoothness parameter S_1 and the smoothness parameter S_2, used to curve fit the biomass concentration and the substrate concentration versus time data, was

$$F = 1.5783 + 0.0120S_1 - 0.0448S_2 - 0.00074S_1^2 + 0.00028S_2^2 - 0.000035S_1S_2 \qquad (3.16)$$

The stationary points of a function (values of the independent variables that set the partial derivatives of a function equal to zero) define either a maximum, minimum or saddle point for that function. Using the methods of differential calculus, the point given by $S_1 = 61$ and $S_2 = 84$ was found to be a stationary point of Eq. (3.16); however, this point is a saddle point. In order to locate the values of S_1 and S_2 which located a minimum on the surface, a direct search procedure, the metohd of successive variation of parameters, was used with Eq. (3.16). The stationary point was used as the starting point in the search over the surface. The results of the search are shown in Table 3.1. Since Eq. (3.16) is only an approximate representation of the actual surface, negative values of F can result using Eq. (3.16) while Eq. (3.4) must be positive. The table indicates that the values of S_1 and S_2 which minimized the absolute value of F were 34 and 84, respectively. These values were used to fit smoothing spline functions to the biomass concentration and substrate concentration versus time data. The curves generated were deemed the best fit to the data since the fermentation studied was described by estimates for the specific growth rate and the specific rate of substrate consumption such that closure of the carbon balance and the available electron balance was optimized.

Figures 3.1 and 3.2 display the smoothing spline curves generated for the biomass concentration and substrate concentration versus time data using methods I and II for selection of the smoothness parameter. Figure 3.1 shows that at hour 6.0, the spline function fit to the data using the S_1 value calculated from method I passed smoothly through the data. The curve fit by method II however, tended to interpolate the data more. The experimental data possessed a large degree of scatter at hour six where fermentor operation was switched from batch to fed batch. Figures 3.3-3.5 show μ, Q_s and η plotted versus time.

Figure 3.3 demonstrates that the estimated values of the specific growth rate follow similar trends when methods I and II are compared. The variation in values is greater for the cross validation method. Figure 3.4 shows that the two methods also give similar values for the specific rate of substrate consumption, Q_s. The first and second values are considerably larger than one would expect for both methods. The differences between the two methods appear to be greater in Figure 3.5 where values of the biomass yield, η, are compared. In Figure 3.5, the variation in the estimates from method II are larger than those from method I. The results from method I show much less variation near the transition point from batch to fed batch operation; for method II variation of the estimates was as large as 24.7% in this region. Smoothness of the estimates of the rate quantities μ, Q_s and η, obtained using method I, indicate that the response surface technique coupled with smoothing spline functions is a suitable procedure for estimation of derivative values. However, the smoothness of the estimates alone does not determine the suitability of the technique. As stated above, the criterion used to select the procedure of method I for calculation of μ, Q_s and η over method II, was closure of the carbon balance and the available electron bal-

ance.

Figures 3.6 and 3.7 show point values for the carbon balance and the available electron balance calculated from Eqs. (3.5) and (3.6) using estimates for the specific growth rate and specific rate of substrate utilization determined from methods I and II. Figures 3.6 and 3.7 shows considerable similarity in the results from the two methods. Using these point values, Eq. (3.4) was evaluated, and F values of 0.0857 and 0.1001 were calculated for the estimates of μ and Q_s determined from methods I and II, respectively. Consequently, the technique for fitting smoothing spline functions to experimental data using a response surface provided values for μ_i and Q_{si} which best closed the carbon balance and available electron balances. Thus, these values appear to be the most accurate estimates of these rate quantities.

For proper identification of the state of a bioreactor precise estimates for the specific growth rate, the specific rate of substrate utilization and the biomass energetic yield are necessary. These quantities are also important in determination of the maximum biomass energetic yield, η_{max}, and maintenance coefficient, m_e. Estimates for η_{max} and m_e were obtained from Pirt's model written in the form of Eqs. (3.15). Point and interval estimates for η_{max} and m_e have been obtained by Solomon *et al.*[9]. The results reported are listed in Table 3.2 along with the values for η_{max} and m_e calculated from the μ_i and Q_{s_i} determined by methods I and II. The estimates determined using the former procedure were designed "RS", to represent the response surface minimized in the method, and those determined using the latter procedure were designed "CV", to represent cross validation. Also listed in Table 3.2, are the data sources used to obtain each estimate. Estimates for the true biomass energetic yield and maintenance coefficient were calculated with and without covariates.

Table 3.2 shows that the method of smoothing the data has an effect on the parameter estimates, η_{max} and m_e. The results with one covariate, η_{max} = 0.603 and m_e = 0.002 hr^{-1} were previously selected as the best estimates by Solomon et al.[10]. For the response surface method of derivative estimation, the parameter estimates with no covariates, η_{max} = 0.570 and m_e = 0.037 hr^{-1}, have the shortest confidence interval for η_{max} and the second shortest confidence interval for m_e. When cross validation was used to estimate the derivatives, only two of the estimates had positive values for the maintenance parameter. The values of η_{max} = 0.637 and m_e = 0.051 hr^{-1}, which are based on the carbon dioxide and biomass measurements, appear to be the cross validation results with shortest confidence interval for which the maintenance parameter estimate is positive. The confidence interval for the maintenance parameter is shortest for this estimate. The estimates of η_{max} of 0.570 and and 0.637 are both within the 95% confidence interval of the estimate of Solomon et al.[10] which is $0.562 \leq \eta_{max} \leq 0.649$. The estimates for the maintenance coefficient of 0.002 hr^{-1} and 0.051 hr^{-1} are within the 95% confidence interval of the response surface estimate which is $-0.022 \leq m_e \leq 0.096$.

The parameter estimates based on the response surface method include only one point estimate in which the maintenance coefficient is negative compared to three estimates using the method of Solomon and four estimates using the cross validation method.

The results in Table 3.2 show that the spline function generated derivative estimates give reasonable parameter estimates, and that the results depend on the method that is used to estimate the derivatives.

Since the variation in the results is fairly large, it is desirable to use consistency tests and other criteria such as prior experience to evaluate the quality of the results which are obtained. Equation (3.4) may be used with both spline function methods to evaluate the consistency of the deriva-

tive estimates.

The use of computers in data collection and analysis is increasing rapidly. Spline functions appear to work reasonably well for smoothing data and estimating derivative quantities.

CONCLUSIONS

A new method of selecting smoothing spline functions, the response surface method and the cross validation method were used and compared for batch followed by fed batch culture data. The newly developed response surface method selects smoothing parameter values which minimize an error function based on the closure of the carbon and available electron balances. Based on this error function the response surface method gave better derivative estimates than the cross validation method. When the estimates of the specific growth rate and specific rate of substrate consumption were compared for the two methods, the estimates are quite similar.

Estimated values of true growth yield and maintenance coefficient depend on the method used to smooth the data. Each of the methods determine reasonable estimates of these parameters.

REFERENCES

1. Reinsch, C. H., "Smoothing by Spline Functions, II", Numerische Mathemtik, 16, 451 (1971).

2. Wold, S., "Spline Functions in Data Analysis", Technometrics, 16, 1, (1974).

3. Wahba, G. and S. Wold, "A Completely Automatic French Curve: Fitting Spline Functions by Cross Validation," Comm. Statist. 4, 1 (1975).

4. Silverman, B. W., "A Fast and Efficient Cross Validation Method for Smoothing Parameter Choice in Spline Regression," J. of the American Statistical Association, 79, 587 (1984).

5. Box, G. E. P. and J. S. Hunter: "Multifactor Experimental Designs for Exploring Response Surfaces," Ann. Math. Stat., 28, 195 (1957).

6. Box, G. E. P. and K. B. Wilson: "On the Experimental Attainment of Optimum Conditions", J. Roy. Statist. Soc., B13, (1951).

7. Box, G. E. P., "The Exploration and Exploitation of Response Surfaces: Some General Considerations and Examples," Biometrics, 10, 16, (1957).

8. Myers, R. H., Response Surface Methodology, R. H. Myers, Virginia Polytechnic Institute, Blacksburg, VA (1976).

9. Solomon, B. O., L. E. Erickson, and S. S. Yang, "Utilization of Statistics and Experimental Design in Data Collection and Analysis," Biotechnol. Bioeng., 25, 2683 (1983).

10. Pirt, S. J., Principles of Microbe and Cell Cultivation, Blackwell, London (1975).

11. Yang, S. S., B. O. Solomon, M. D. Oner, and L. E. Erickson, "Estimating and Testing Common Parameters for Some Multiresponse Models Associated with Microbial Growth and Bioenergetics," Technometrics, 20, 355.

12. SAS Users Guide, SAS Institute, Raleigh, N. C. (1979).

NOMENCLATURE

Roman Letters

b_0, b_1, b_2, b_{11}, b_{22}, b_{12}	=	coefficients of response surface fitted to the data
D	=	dilution rate the $i^{\underline{th}}$ data point
d_{yi}	=	weight of $i^{\underline{th}}$ data point
F	=	sum of the squares of the deviations of the carbon and available electron balances from one
g	=	smoothing spline function
G	=	integral of the square of the second derivative of the smoothing spline function
g_s^{-1}	=	spline function fitted to N-1 data points with S as the value of the smoothness parameter
J_a	=	available electron balance
J_{ai}	=	available electron balance at time t_i
J_c	=	carbon balance
J_{ci}	=	carbon balance at time t_i
m_e	=	maintenance coefficient
N	=	number of data points
Q_{ci}	=	specific rate of carbon dioxide evolution at t_i
Q_{oi}	=	specific rate of oxygen consumption at t_i
Q_{si}	=	specific rate of substrate consumption at t_i
S	=	smoothness parameter
S	=	substrate concentration
S_f	=	inlet substrate concentration
S_j	=	$j^{\underline{th}}$ value of the smoothness parameter
S_1	=	smoothness parameter used to fitted spline function to biomass concentration data
S_2	=	smoothness parameter used to fitted spline function to substrate concentration data

t = time

X = biomass concentration

Greek Letters

α_1 = coded value for the upper bound of S_1

$-\alpha_1$ = coded value for the lower bound of S_1

α_2 = coded value for the upper bound of S_2

$-\alpha_2$ = coded value for the lower bound of S_2

γ_b = reductance degree of biomass

γ_s = reductance degree of substrate

η = biomass energetic yield

η_{max} = true growth yield

μ = specific growth rate

σ_b = weight fraction carbon in biomass

σ_s = weight fraction carbon in substrate

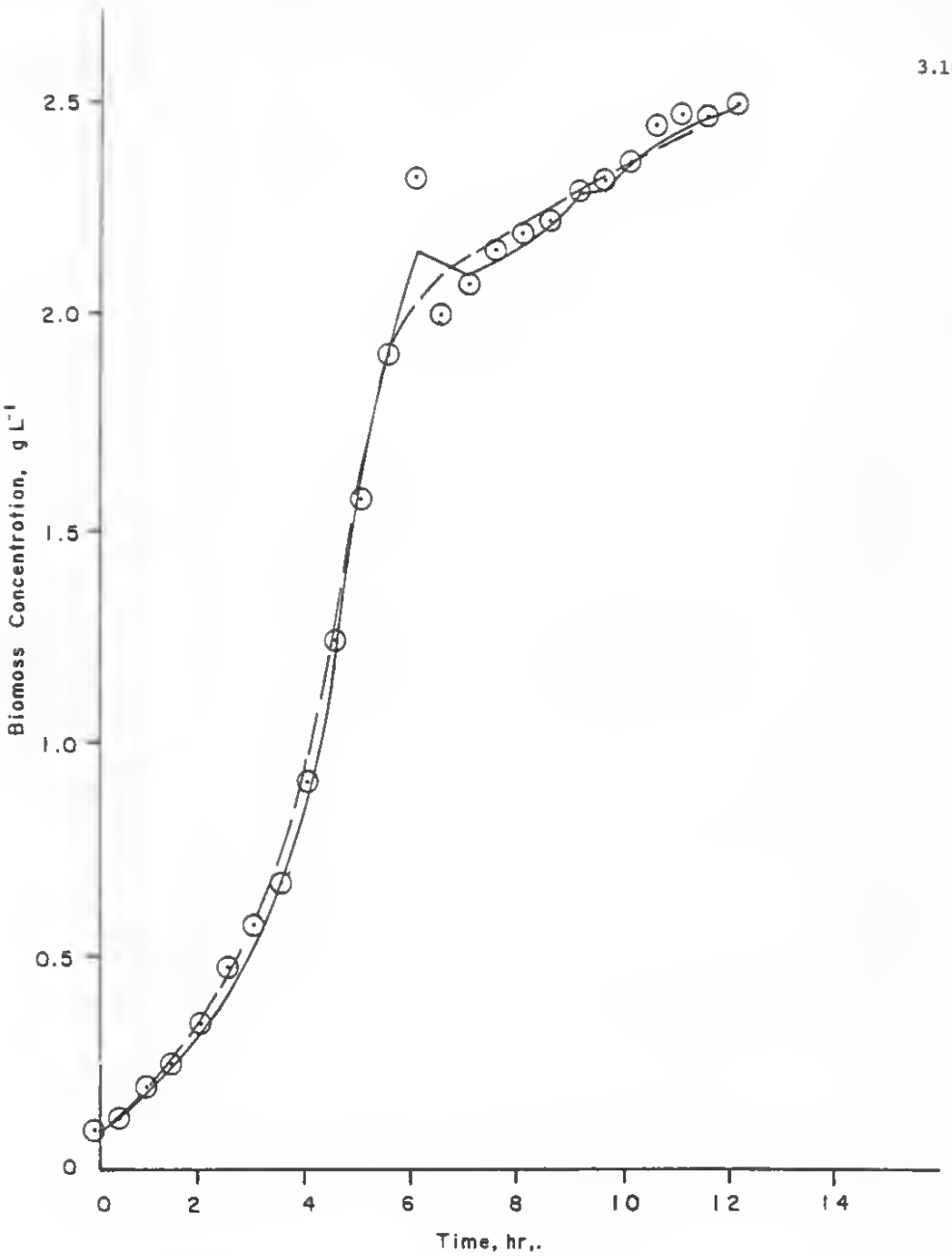


Figura 3.1 Smoothing spline curvas fittad to biomss concantration data for a batch/fed-batch fermentation of Solomon at al. (1983). The daehed line repreeents the spline function generated for S=61 as calculeted from method I. Tha solid lina raprasante eplina function ganaratad from mathod II. The circles rapresant the experimantal data.

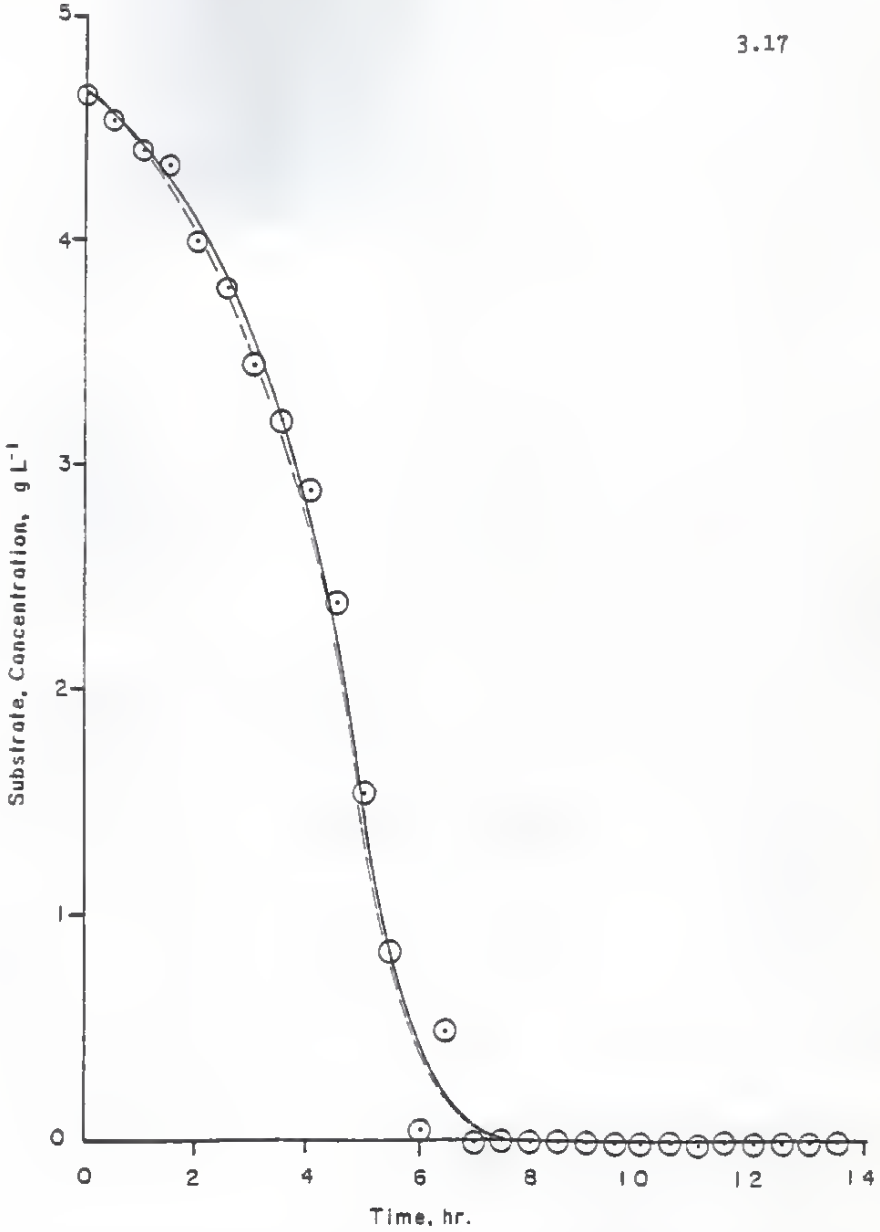


Figure 3.2 Smoothing spline curves fitted to substrate concentration data for a batch/fed batch fermentation. The dashed line represente the epline function generated for S=84 as calculated from method I. The solid line repreeents the epline function generated from method II. The circlee repreeent the experimental data.

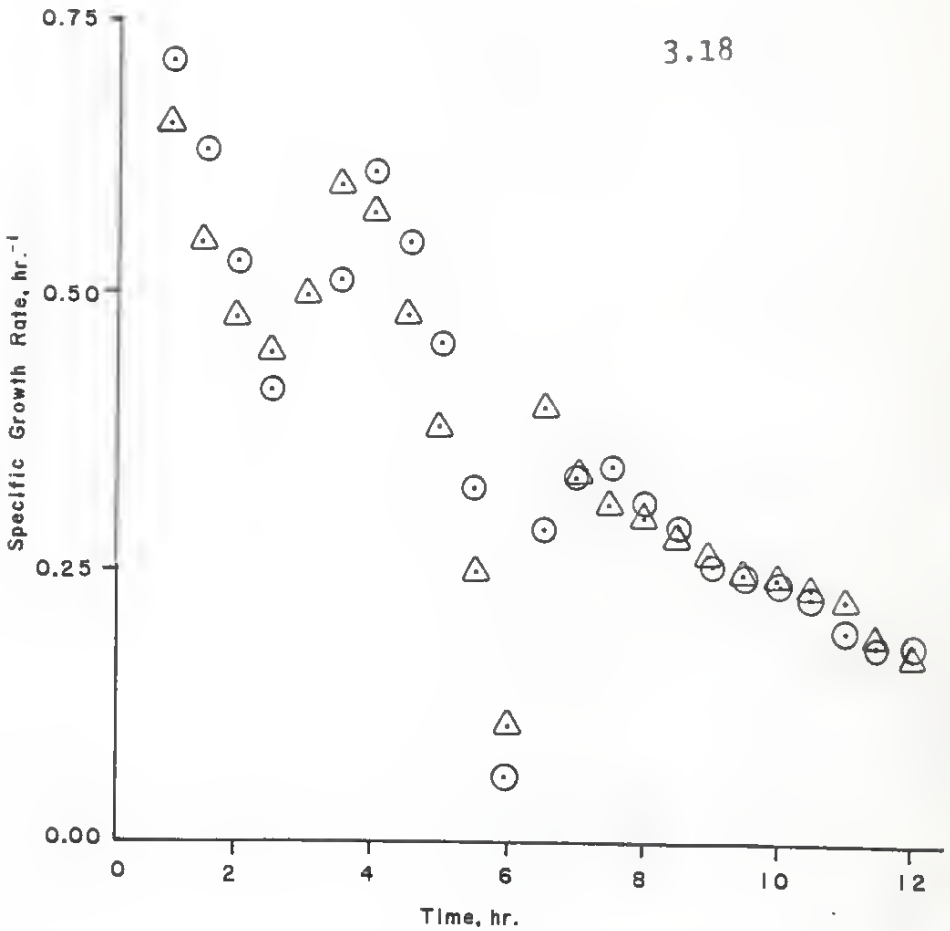


Figure 3.3 Estimated values of specific growth rate for a batch/fed-batch fermentation. The triangles represent the estimates obtained using method I, and the circles represent the estimates obtained using method II.

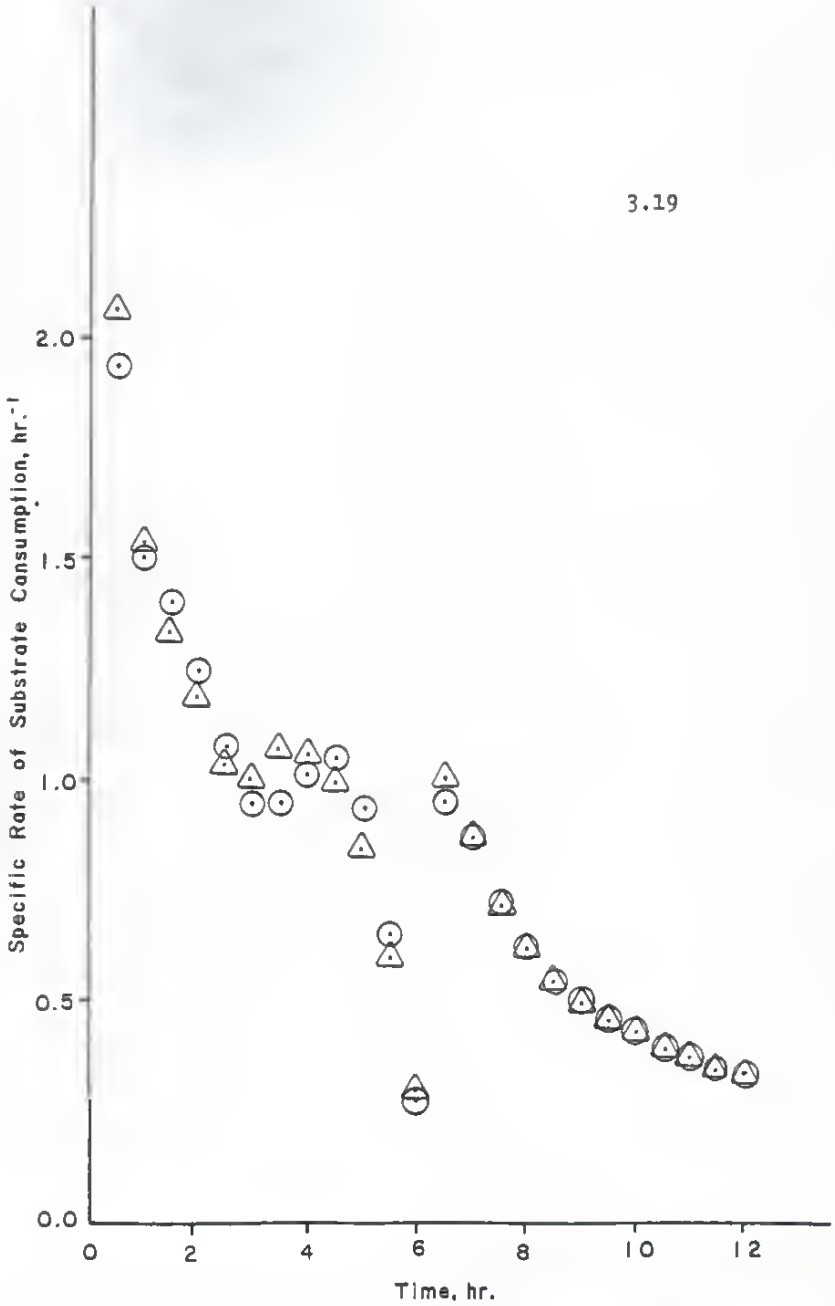


Figure 3.4 Estimated values of the specific rate of subatrate consumption for a batch/fed-batch fermentation. The triangles represent the estimates obtained using method I, and the circles represent the estimates obtained using method II.

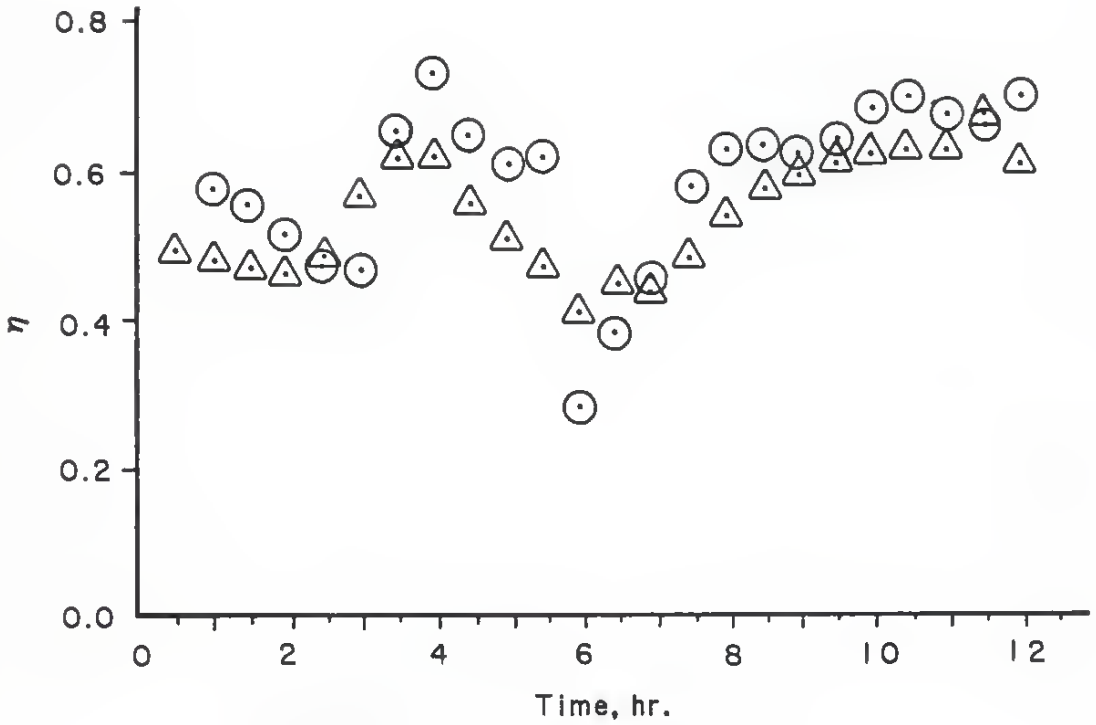


Figure 3.5 Estimates values of the biomass energetic yield, η, for a batch/fed-batch fermentation. The triangles represent the estimates obtained using method I, and the circles represent the estimates obtained using method II.

Figure 3.6 Plot of the carbon balance versus time. The triangles represent the balance calculated from the specific rate estimates made using method I, and the circles represent the balance calculated from the specific rate estimates made using method II.

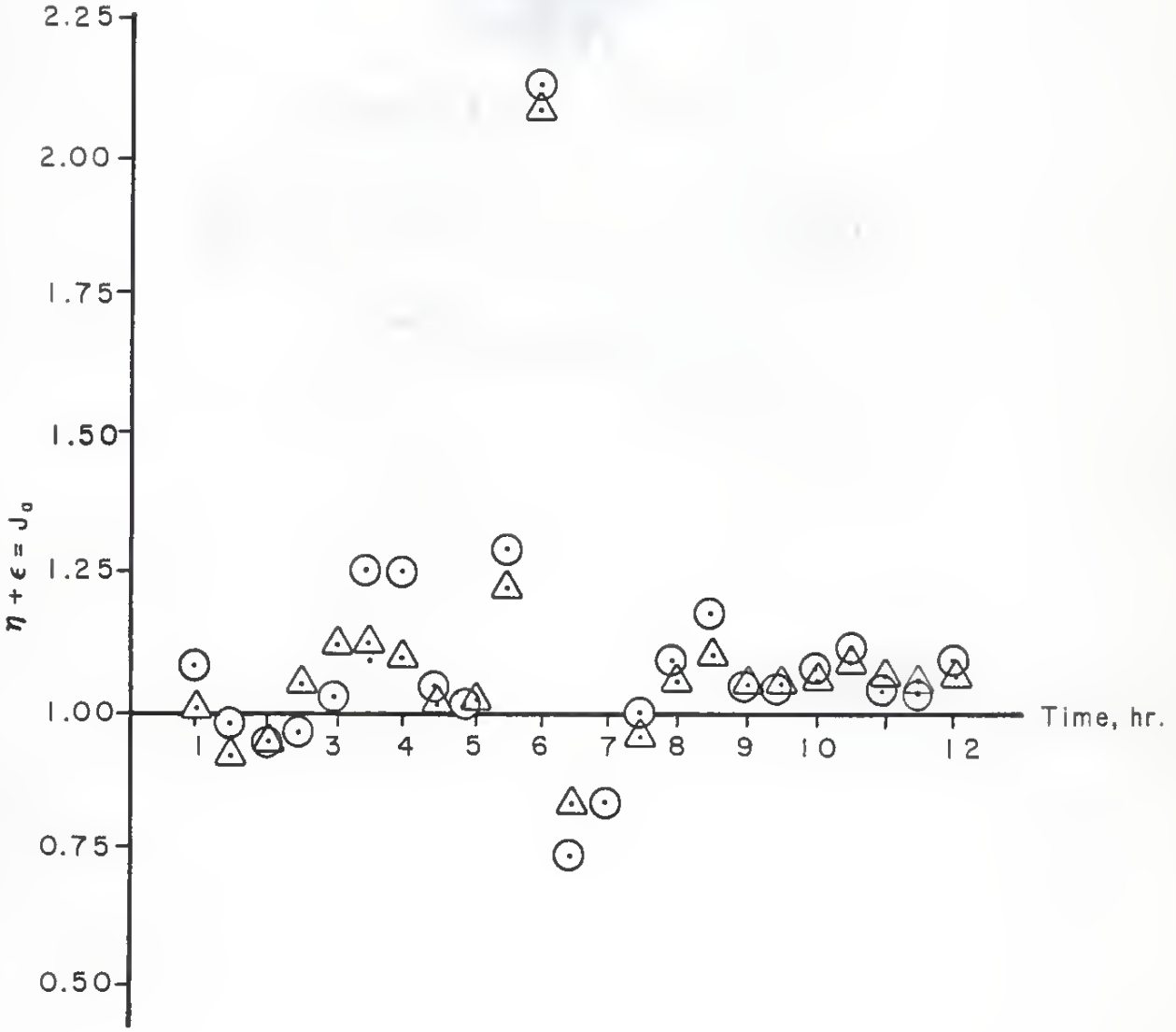


Figure 3.7 Plot of the available electron balance versus time. The triangles represent the balance calculated from the specific rate estimates made using method I, and the circles represent the balance calculated from the specific rate estimates made using method II.

Table 3.1. Results for the Optimization of the Response Surface of Eq. (3.16) Using the Method of Successive Variation of Parameters.

S_1	S_2	F
61	84	0.05528
61	88	0.05850
61	86	0.05578
61	83	0.05586
50	84	0.04556
40	84	0.02109
20	84	-0.07252
30	84	-0.01800
35	84	0.00327
36	84	0.00713
34	84	-0.00074

Table 3.2. Estimates of the True Biomass Energetic Yield and Maintenance Coefficient Determined Using Three Procedures for Calculation of the Specific Growth Rate and the Specific Rate of Sub-Strate Utilization.

		True Growth Yield, η_{max}					
Data	Covariates Included	Point(S)	Point(RS)	Point(CV)	CI(S*)	CI(RS)	CI(CV)
Q_s, μ	none	0.567	0.498	0.450	[0.492, 0.668]	[0.459, 0.544]	[0.388, 0.535]
Q_o, μ	none	0.532	0.547	0.479	[0.494, 0.577]	[0.501, 0.602]	[0.442, 0.522]
Q_c, μ	none	0.653	0.699	0.637	[0.607, 0.708]	[0.648, 0.758]	[0.587, 0.697]
All	Z_1, Z_2	0.612	0.724	0.637	[0.548, 0.694]	[0.584, 0.953]	[0.536, 0.786]
All	Z_1	0.603	0.569	0.529	[0.563, 0.649]	[0.527, 0.619]	[0.489, 0.575]
All	none	0.580	0.570	0.510	[0.532, 0.637]	[0.531, 0.614]	[0.464, 0.568]

		Maintenance Coefficient, m_e, $hr.^{-1}$					
Data	Covariates Included	Points(S)	Point(RS)	Point(CV)	CI(S)	CI(RS)	CI(CV)
Q_s, μ	none	-0.029	-0.050	-0.137	[-0.100, 0.042]	[-0.129, 0.029]	[-0.292, 0.030]
Q_o, μ	none	-0.034	0.051	-0.059	[-0.073, 0.005]	[-0.027, 0.129]	[-0.136, 0.019]
Q_c, μ	none	0.020	0.110	0.051	[-0.011, 0.051]	[0.057, 0.162]	[-0.010, 0.112]
All	Z_1, Z_2	0.006	0.117	0.059	[-0.032, 0.044]	[0.027, 0.207]	[-0.024, 0.144]
All	Z_1	0.002	0.036	-0.048	[-0.028, 0.033]	[-0.023, 0.095]	[-0.079, 0.059]
All	none	-0.014	0.037	-0.048	[-0.055, 0.027]	[-0.022, 0.096]	[-0.138, 0.041]

* S = Solomon et al.[9]; RS = Response Surface; CV = Cross Validation; CI = 95% Confidence Interval.

CHAPTER IV

CURVE FITTING EXPERIMENTAL BATCH/FED BATCH FERMENTATION DATA USING SMOOTHING SPLINE FUNCTIONS

INTRODUCTION

A systematic procedure to curve fit the experimental fermentation data of Sclomon et al.[1] and to estimate rate quantities, which utilizes smoothing spline functions, is described in Chapters 2 and 3. The procedure for smoothing biomass concentration and substrate concentration data collected from a batch followed by a fed-batch fermentation is illustrated.

An n^{th} order spline function is a continuous function with n-1 continuous derivatives. However, when a fermentation is switched from batch operation to fed-batch operation by starting the addition of feed at a constant feeding rate, the step change in feed flow rate results in a step change in the first derivatives of the biomass and substrate concentrations. Thus, one should not expect these derivative quantities to be continuous when there is a step change in the feed flow rate. Separate spline functions may be fit to the batch data and fed-batch data; this is a solution which can be appropriately considered when the number of step changes in feeding rate are small. Since good derivative estimates are not usually obtained at the end points, some information is lost at the point of discontinuity.

The purpose of this work was to examine the consistency of specific rate estimates from smoothing spline functions fit separately to the batch and fed-batch portions of the data collected from a batch followed by fed-batch fermentation. In addition, the true growth yield and the maintenance parameter were estimated and compared with earlier estimates.

THEORY

Curve Fitting the Data

The mathematics of smoothing spline functions is well developed.[2-6] The smoothing parameters, S_1 and S_2, may be selected using a response sur-

face method. The spline function IMSL subroutine ICSSCU minimizes

$$G = \int_{t_o}^{t_N} \left(\frac{d^2 g}{dt^2}\right)^2 dt \tag{4.1}$$

subject to

$$S_j \geq \sum_{i=1}^{N} \frac{[y_i - g(t_i)]^2}{(dy_i)^2} \tag{4.2}$$

where g(t) is the spline function for the biomass data with $y_i = X_i$ for the smoothing parameter $S_j = S_1$ and similarly g(t) is the spline function and y_i is the substrate data for the smoothing parameter $S_j = S_2$. The values of dy_i, i = 1, 2, ..., N, are input parameters which enable the user to weigh the data. A response surface method[7] is used to select S_1 and S_2 to minimize

$$F = \frac{1}{N} \sum_{i=1}^{N} [(J_{ci} - 1)^2 + (J_{ai} - 1)^2] \tag{4.3}$$

where

$$J_{ci} = \frac{\frac{\mu_i \sigma_b}{12} + Q_{ci}}{\frac{\sigma_s Q_{si}}{12}} \tag{4.4}$$

$$J_{ai} = \frac{\frac{\mu_i \sigma_b \gamma_b}{12} + 4Q_{oi}}{\frac{\sigma_s \gamma_s Q_{si}}{12}} \tag{4.5}$$

The response surface method is Method I in this work. Method II is the cross validation method[3,6] which is carried out using the IMSL subroutine ICSSCV. The cross validation method selects an optimal value of the smoothing parameter, S, which minimizes

$$XVSC = \frac{1}{N} \sum_{i=1}^{N} [y_i - g_s(t_i)]^2 \qquad (4.6)$$

where $g_s(t_i)$ is the spline function for value S of the smoothing parameter and point i deleted from the data set. With this method each curve is fit independently.

Parameter Estimation

For the case of no product formation, Pirt's model of microbial growth states that the available electrons present in the substrate are allocated for growth and for maintenance, viz.,

$$\frac{\mu}{\eta} = \frac{\mu}{\eta_{max}} + m_e \qquad (4.7)$$

Estimates for the specific growth rate, μ, and the biomass energetic yield, η, obtained from curves fitted to the batch data and from curves fitted to the fed-batch data are combined to give one complete set of estimates for μ and η which describe the entire fermentation. Two complete sets of estimates are obtained using methods I and II. From the two sets of estimates, two point estimates for η_{max} and m_e are calculated from Eq. (4.7). The carbon and available electron balances can be combined with Eq. (4.7) to give additional equations for point estimation of η_{max} and m_e, viz.,

$$\frac{\mu}{\eta} (y_c + d) = \frac{\mu}{\eta_{max}} + m_e \qquad (4.8)$$

$$\frac{\mu}{\eta} (\varepsilon + \eta) = \frac{\mu}{\eta_{max}} + m_e \qquad (4.9)$$

Equations (4.8) and (4.9) introduce into the estimation scheme CO_2 and O_2 measurements which are not used in the point estimates made using Eq. (4.7). The covariate adjustment method[1,8,9] is used to obtain combined estimates

using all of the data simultaneously.

MATERIALS AND METHODS

The batch/fed-batch data of 5-3-82 of Solomon and co-workers[1] were used in this study. The batch portion of the data from hours 0 to 6.0 was treated separately from the fed-batch portion of the data for hours 6.5 to 13.5. For method I, the IMSL subroutine ICSSCU was used with SAS[10] to determine the response function, Eq. (4.3), for values of the smoothing parameters S_1 and S_2. For method II, the IMSL subroutine ICSSCV was used to fitted spline functions to the data by using cross validation to select the smoothness parameters.

All point estimates and confidence intervals for η_{max} and m_e were determined by linear regression using SAS[10].

RESULTS AND DISCUSSION

Batch Fermentation

The response surface generated for the data collected from hours 0.0 to 6.0 is,

$$F = 0.1019 - 0.00091S_1 - 0.0091S_2 + 0.000047S_1^2 + 0.0027S_2^2 - 0.000071S_1S_2 \qquad (4.10)$$

A stationary point for Eq. (4.10) was calculated by taking partial derivatives of F with respect to S_1 and S_2, setting the resulting equations equal to zero and solving the equations simultaneously. The point $S_1 = 18.45$ and $S_2 = 23.51$ is a minimum for the surface of Eq. (4.10). The spline functions generated for biomass concentration and substrate concentration data using $S_1 = 18.45$ and $S_2 = 23.51$ are shown in Figures 4.1 and 4.2. The smooth curves yield

correspondingly smooth estimates for the specific growth rate and the specific rate of substrate utilization. This is shown in Figures 4.3 and 4.4. Estimates made for the biomass energetic yield calculated from the specific rate quantities are relatively constant throughout the time interval from 0.0 to 6.0 hrs. However, towards the end of the fermentation, as expected, specific growth rate decreases. This results in a corresponding decrease in estimates for the biomass yield. The biomass yield is plotted in Figure 4.5. Scatter present in the estimates for the specific rate of substrate consumption, obtained using method II, introduced scatter into the estimates for the biomass energetic yield.

Figures 4.6 and 4.7 display values of the carbon balance and the available electron balance plotted versus time. The values were calculated using estimates for specific growth rate and specific rate of substrate utilization calculated from method I and method II. F values were calculated for methods I and II, and were 0.0201 and 0.0966, respectively. The F values show that the data calculated from spline functions fit using method I are more consistent than the estimates determined using method II.

Fed-Batch Fermentation

Figures 4.8 and 4.9 display the spline functions fit to the biomass concentration and substrate concentration fed-batch data using methods I and II. The response surface calculated was,

$$F = 0.0479 + 0.00022S_1 - 0.0022S_2 - 0.0000058S_1^2 + 0.000039S_2^2 - 1.7 \times 10^{-20}S_1S_2 \qquad (4.11)$$

To determine values for S_1 and S_2 which minimize Eq. (4.11), the stationary point for Eq. (4.11) was calculated. The point given by $S_1 = 19.21$

and S_2 = 27.97 was the stationary point located. From differential calculus it was determined that the stationary point defines a saddle point on the surface described by Eq. (4.11). A quasi Newton method was used to determine the values for S_1 and S_2 which minimize Eq. (4.11). The stationary point was selected as the first guess for the search on S_1 and S_2. A minimum was located at S_1 = 18.97 and S_2 = 28.01. From the curves generated, the specific growth rate and the specific rate of substrate utilization were estimated. The specific rates are plotted versus time in Figures 4.10 - 4.11. These figures indicate that the specific growth rate and specific rate of substrate utilization estimates calculated from method I trace somewhat smoother curves than the estimates determined from method II. This is evidenced by the plot of the biomass energetic yield shown in Figure 4.12. The rate estimates determined by methods I and II were examined for consistency by evaluation of the carbon balance and available electron balance. An F value of 0.0166 was calculated from the results for method I. For method II, large values for substrate uptake coupled with relatively constant estimates for the microbial growth rate gave low initial values for the calculated biomass yields. As a consequence, point values for the carbon and available electron balances are quite low at the outset of the fermentation compared to the values calculated from method I. The balances are plotted versus time in Figures 4.15 and 4.14. An F value of 0.0472 was calculated from the results for method II. The rate estimates obtained from method I are more consistent than those obtained from method II. Based on the consistency of the estimates, the former estimates offer a better description of the state of the bioreactor.

Three sets of estimates for the true biomess energetic yield and maintenance coefficient are presented in Table 4.1. The estimates were determined from three sets of estimates for the specific growth rate and the

specific rate of substrate utilization made using three procedures to curve fit experimental data. Each of the three sets of estimates consist of six point and interval estimates.

The estimates obtained previously by Solomon et al.[1] are identified with an "S"; the results obtained using the response surface method of smoothing (method I) are denoted by "RS"; and "CV" refers to cross validation (method II). The measurements associated with Eq. (4.7) are Q_s and μ; Q_o and μ are used with Equation (4.9); and Q_c and μ are used with Eq. (4.8). The other estimates are obtained using the covariate adjustment method[1,8,9] and all of the data.

The best estimates for each set were selected by examination of the width of the 95% confidence interval. In the regression, estimates for specific rates at t = 0.0, which are based on one sided derivatives are deleted from the analysis. Point values at t = 6.5 where fermentor operation had been switched from batch to fed-batch are also deleted. In addition, the data at t = 0.5 were also not used, for method I, due to unreasonably large values for specific growth rate. The point estimate of 0.594 for η_{max} made from all of the data using one covariate has the smallest 95% confidence interval. The width of the interval estimator is .097. The corresponding point estimate for m_e is 0.013 $hr.^{-1}$ which has the second smallest confidence interval of the method I estimates for m_e. The estimate for m_e = -0.008 $hr.^{-1}$ has the smallest 95% confidence interval, but it is rejected due to the negative value for m_e and the larger corresponding interval for η_{max}. For method II, the point estimate η_{max} = 0.585 made from all data with one covariate has the second smallest interval estimator, i.e., 0.096. The estimate for m_e corresponding to η_{max} = 0.585 is the 0.008 $hr.^{-1}$ which has the third smallest confidence interval. The smallest intervals for method II are for m_e = -0.012 $hr.^{-1}$ and η_{max} = 0.549 determined from biomass concentration and

oxygen measurement; however, the negative estimate for m_e is not reasonable.

The results may also be compared with the estimates presented in Table 4.2 which were generated from spline functions fit to all the data simultaneously.

The average values of the five selected point estimates in Tables 4.1 and 4.2 are 0.598 for η_{max} and 0.022 for m_e. These average values are included in the 95% confidence intervals for all of the selected results.

Negative values of the maintenance coefficient were observed for both of the methods. For method II, five of six of the estimates are positive in Table 4.1 compared to two of six in Table 4.2

In Table 4.3 the values of F are compared for methods I and II, and for fitting the batch and fed-batch data separately and as one set of data. Both methods I and II satisfy the consistency test better when the data are analyzed as two separate sets of data. The most consistent estimates for the specific rates result from using method I.

The response surface method and cross validation appear to be appropriate methods to use when smoothing fermentation data. An upper limit on the values of the smoothing parameters may be needed in some cases with the response surface method. Forcing the curves to pass through the first and last data points may also be desirable; this can be done by setting dy_i to zero for the first and last data points. Some results where this has been done are presented elsewhere.[11]

Figures 4.15 and 4.16 display values for the carbon balance and available electron balance calculated from methods I and II for the case of fitting spline functions to all the data. Figures 4.6, 4.7, 4.13, 4.14, 4.15, and 4.16 show that considerable improvement in data consistency is observed for the case of fitting spline functions to batch and fed-batch data separately. Theoretically,

this is understandable as the mathematics which describe each mode of fermentor operation define continuous functions and derivatives for biomass and substrate concentration in the fermentor since spline functions are by definition continuous functions, modelling only the continuous parts of the process yields better results.

CONCLUSIONS

The F values obtained using method I applied to the batch and fed-batch data were 0.0201 and 0.0166, respectively. For method II, the F values obtained were 0.0966 and 0.0472 for the batch and fed-batch portion of the data, respectively. These F values are considerably smaller than the F values obtained for methods I and II of 0.0857 and 0.1001, respectively, for the case of splines functions fit to all the data. Consequently, specific rate estimates for the case of spline functions fitted to all the data are less consistent than are the specific rate estimates made from curves fitted separately to the batch and fed-batch portions of the data.

The estimates of the true growth yield, η_{max}, and maintenance coefficient, m_e, depend on the method of smoothing of the experimental data; however, for all of the selected results, the 95% confidence interval included the average value of the point estimates. The response surface method and cross validation method appear to be potentially useful methods for smoothing fermentation data.

REFERENCES

1. Solomon, B. O., L. E. Erickson, and S. S. Yang, "Utilization of Statistics and Experimental Design in Data Collection and Analysis," Biotechnol. Bioeng., 25, 2683 (1983).

2. Wold, S., "Spline Functions in Data Analysis," Technometrics, 16, 1 (1974).

3. Wahba, G., and S. Wold, "A Completely Automatic French Curve: Fitting Spline Functions by Cross Validation," Communications in Statistics, 4, 1 (1975).

4. Reinsch, C. H., "Smoothing by Spline Functions," Numerische Mathematik, 10, 177 (1967).

5. Reinsch, C. H., "Smoothing by Spline Functions. II," Numerische Mathematik, 16, 451 (1971).

6. Silverman, B. W., "A Fast and Efficient Cross Validation Method for Smoothing Parameter Choice in Spline Regression," J. of the American Statistical Association, 79, (1984).

7. Myers, R. H., Response Surface Methodology, R. H. Myers, Virginia Polytechnic Institute, Blacksburg, VA (1976).

8. Yang, S. S., B. O. Solomon, M. D. Oner, and L. E. Erickson, "Estimating and Testing Common Parameters for Some Multiresponse Models Associated with Microbial Growth and Bioenergetics," Technometrics, 20, 355 (1984).

9. Solomon, B. O., M. D. Oner, L. E. Erickson, and S. S. Yang, "Estimation of Parameters Where Dependent Observations are Related by Equality by Equality Constraints," AIChE Journal, 30, 747 (1984).

10. SAS User's Guide, SAS Institute, Raleigh, N.C. (1979).

11. Oner, M. D., "Estimation of Yield and Kinetic Parameters in Lactic Acid Fermentations for Yoghurt Cultures," Ph.D. Dissertation, Kansas State University, Manhattan, Kansas (1985).

NOMENCLATURE

Roman Letters

d	= fraction of substrate carbon evolved as carbon dioxide
$\frac{1}{dy_i}$	= relative weight of the i^{th} data point
F	= sum of the square of the difference of carbon and available election balances from one
$g(t)$	= smoothing spline function
L	= width of 95% confidence interval
m_e	= maintenance coefficient
P	= function minimized to generate smoothing spline functions
Q_s	= specific rate of substrate consumption
S	= smoothing parameter
S_1	= smoothing parameter for curve fitted to biomass concentration data
S_2	= smoothing parameter for curve fitted to substrate concentration data
t	= time
y_e	= fraction of organic substrate carbon incorporated into biomass
y_i	= i^{th} observation

Greek Letters

ε	= fraction of organic substrate available elections transferred to oxygen
η	= biomass energetic yield
μ_{max}	= true growth yield
λ	= smoothing parameter
μ	= specific growth rate

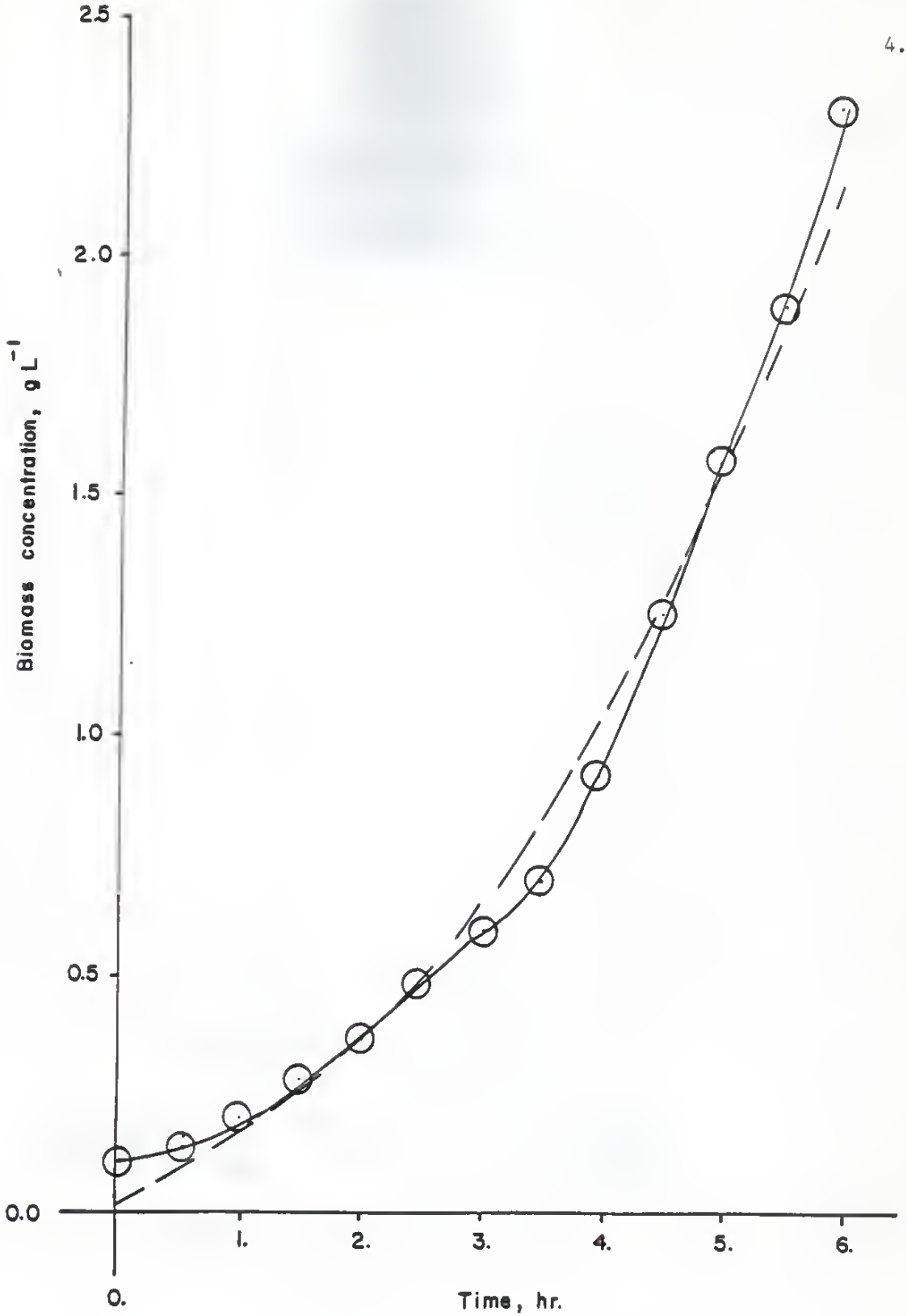


Figure 4.1 Smoothing spline functions fitted to biomass concentration versus time data for the batch fermentation data of Solomon et al. (1983). The dashed line represents the spline function fitted using method I. The solid line represents spline function fitted using method II.

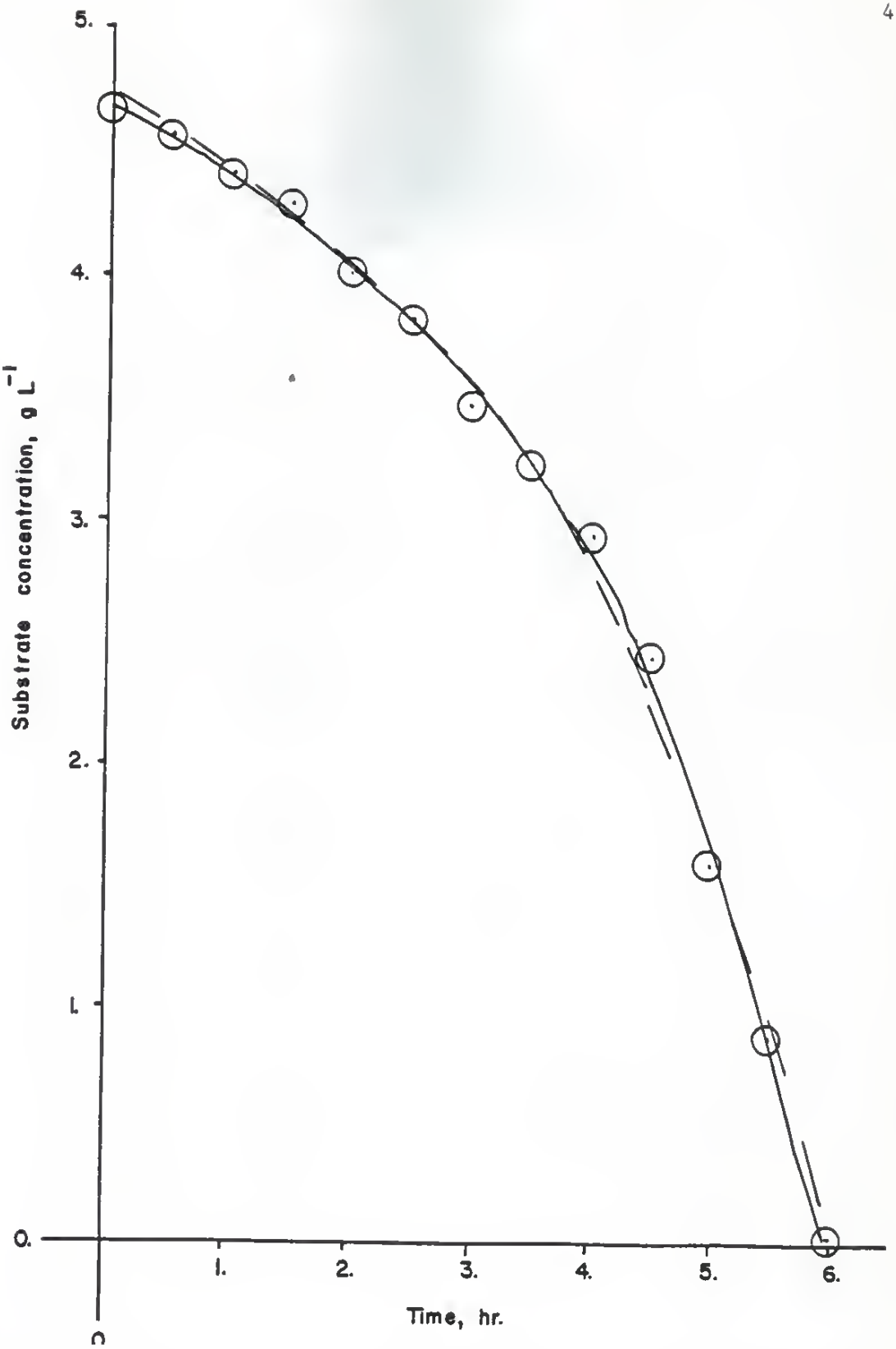


Figure 4.2 Smoothing spline functions fitted to substrate concentration versus time data for the batch fermentation data of Solomon et al. (1983). The dashed line represents spline function fitted using method I. The solid line represents the spline function fitted using method II.

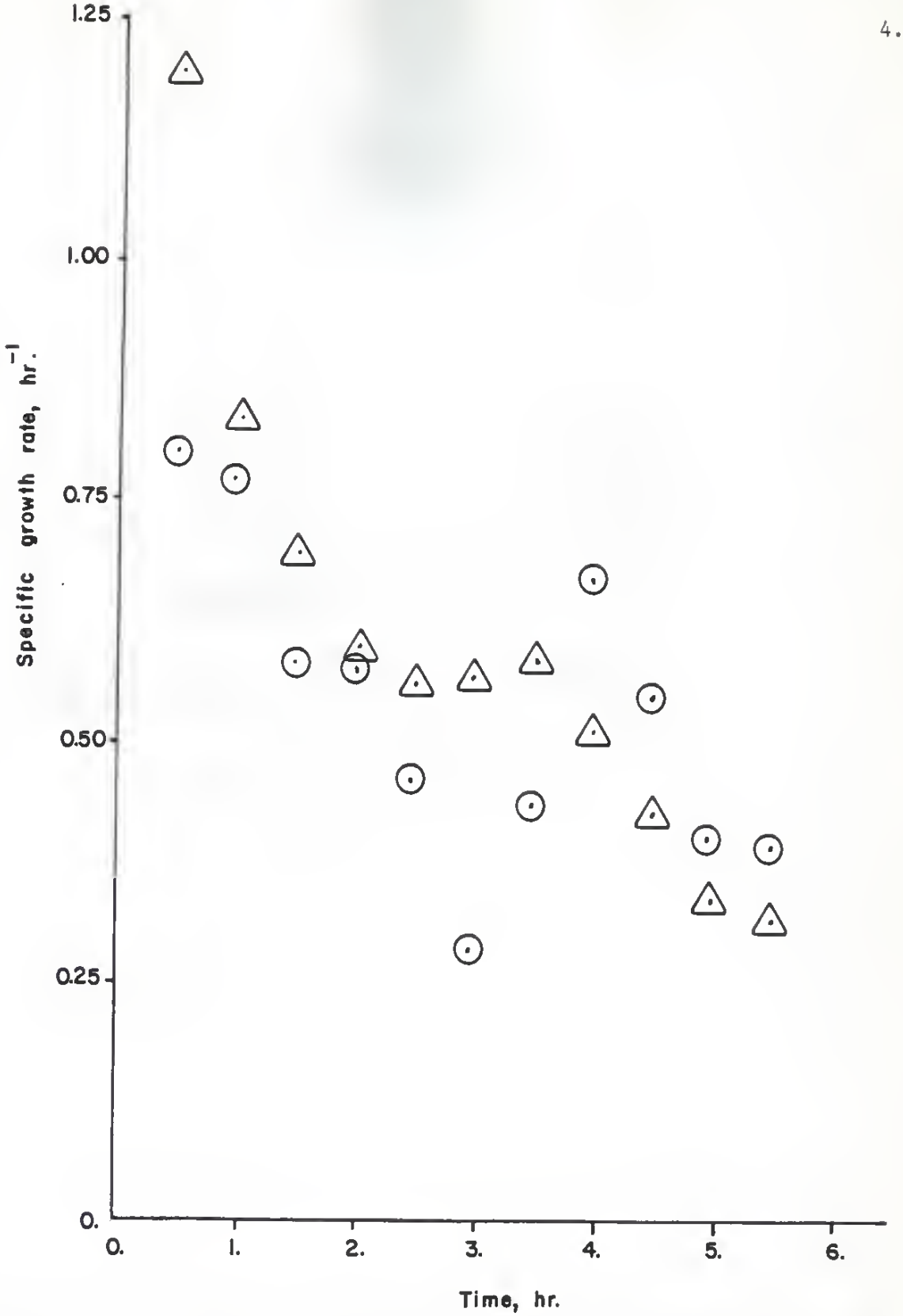


Figure 4.3 Estimates for the specific rate of microbial growth for the batch data. The triangles represent estimates obtained using method I. The circles represent estimates obtained using method II.

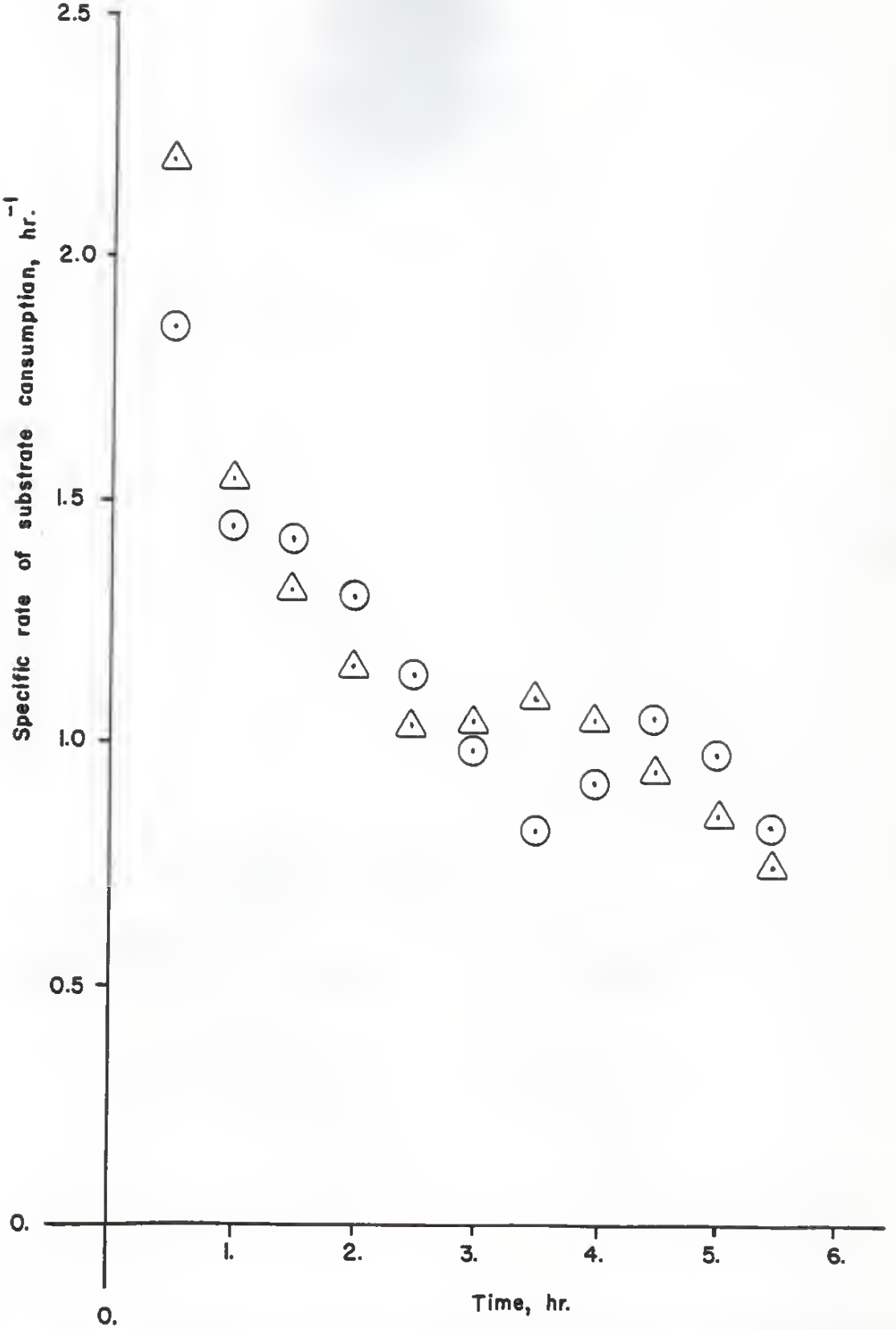
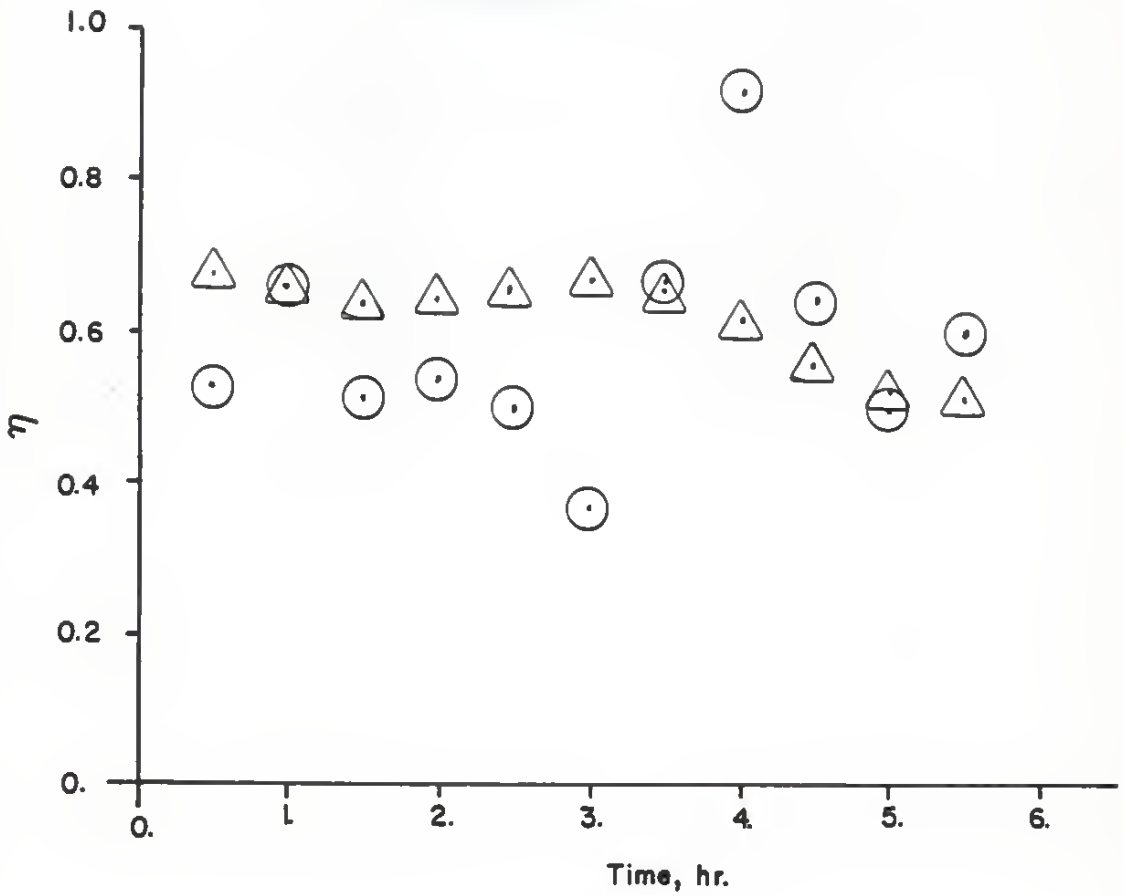


Figure 4.4 Estimates for the specific rate of substrate utilization for the batch data. The triangles represent estimates obtained using method I. The circles represent estimates obtained using method II.



Fibure 4.5 Estimates for the biomass energetic yield for the batch data. The triangles represent estimates obtained using method I. The circles represent estimates obtained using method II.

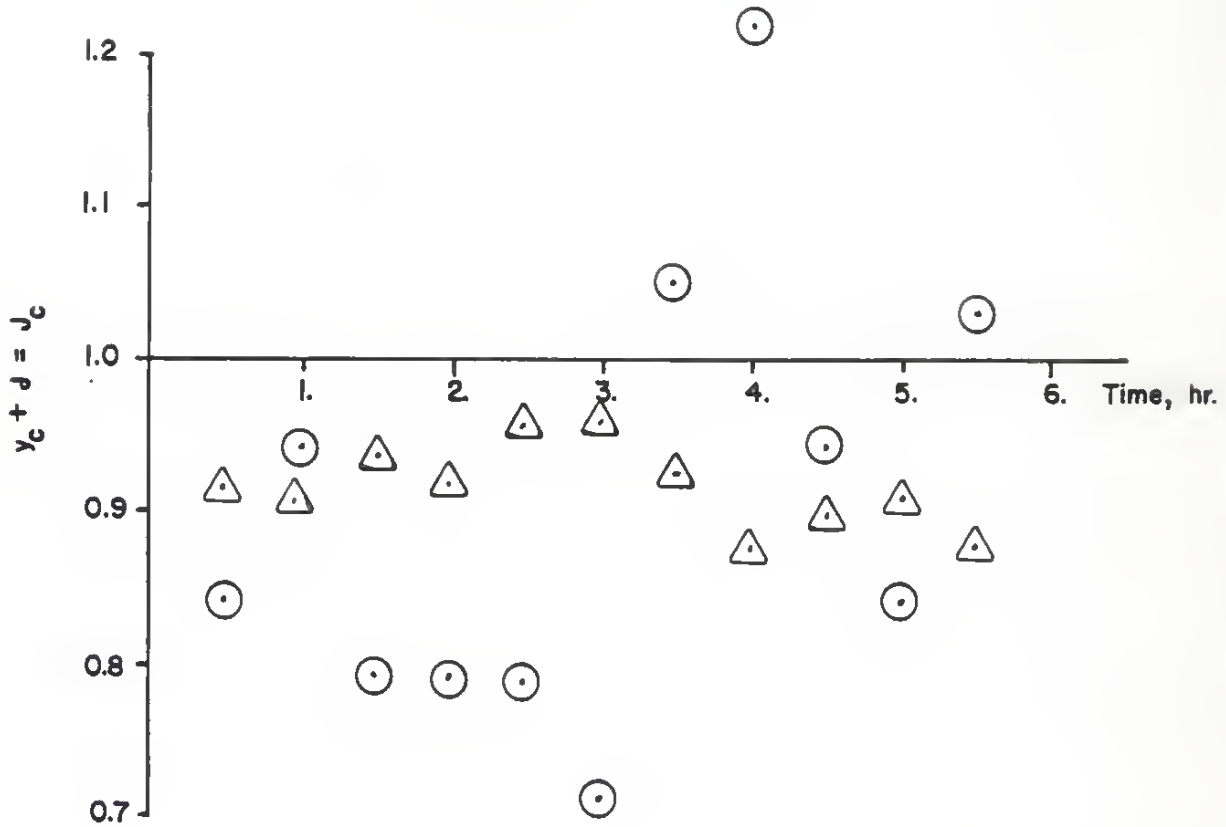


Figure 4.6 Plot of the carbon balance versus time. The triangles represent the balance calculated from specific rate estimates made using method I, and the circles represent the balance calculated from specific rate estimates made using method II.

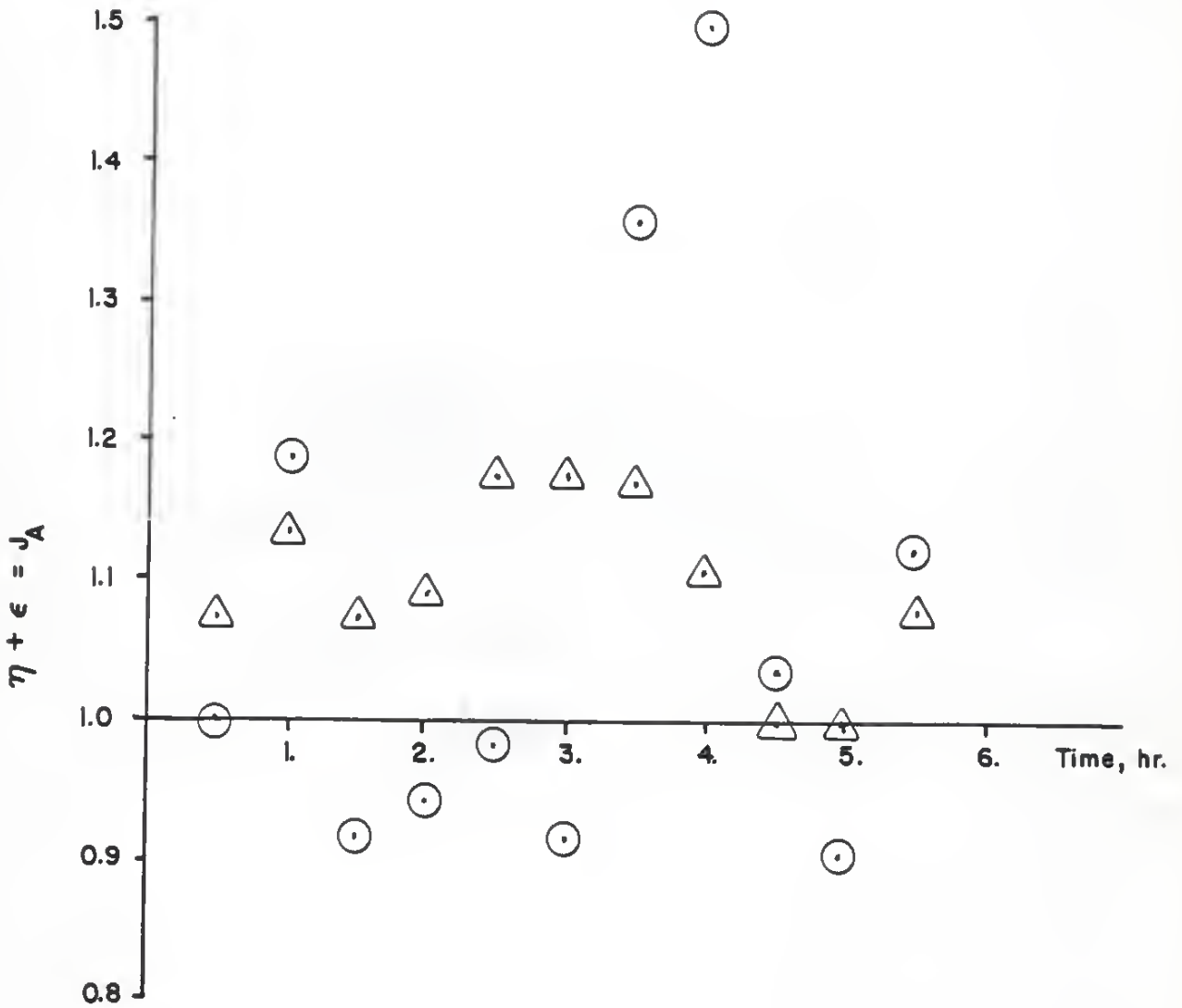


Figure 4.7 Plot of the available electron balance versus time. The triangles represent the balance calculated from specific rate estimates made using method I, and the circles represent the balance calculated from specific rate estimates made using method II.

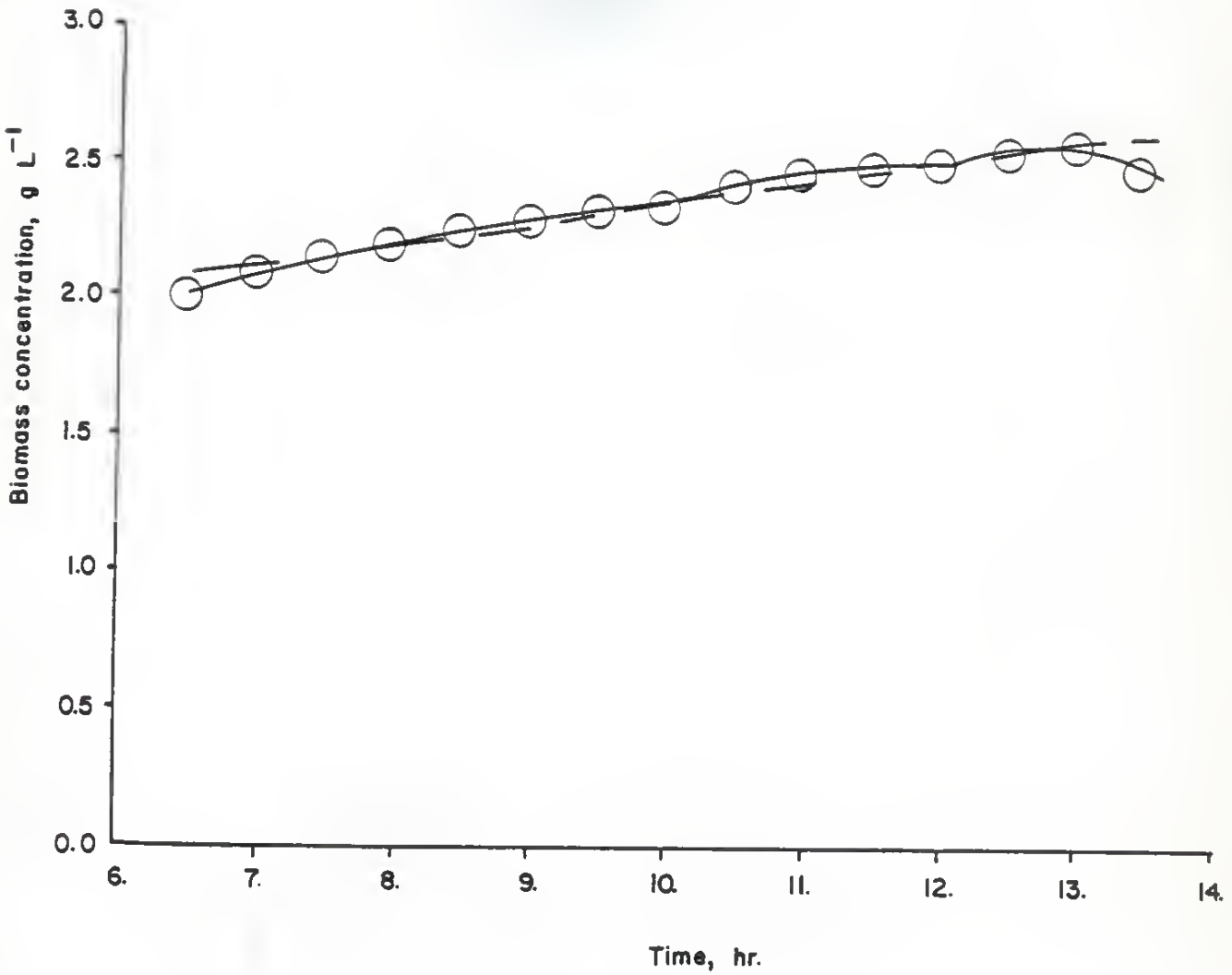


Figure 4.8 Smoothing spline functions fitted to biomass concentration versus time data for the fed batch fermentation data of Solomon, et al. (1983). The dashed line represents the spline function fit using method I. The solid line represents the spline function fitted using method II.

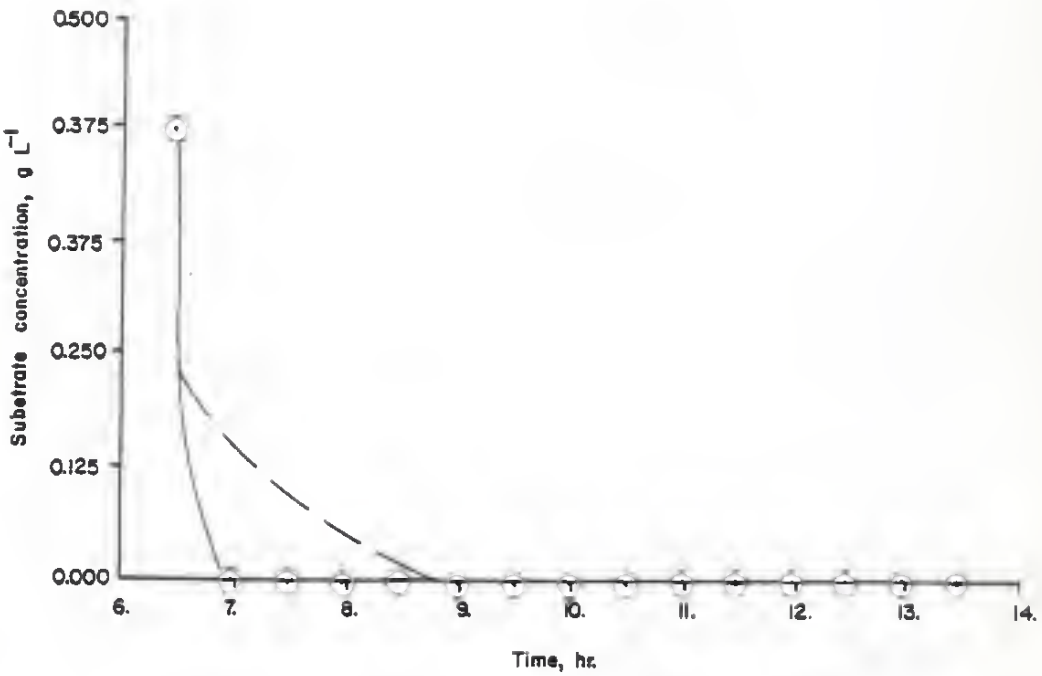


Figure 4.9 Smoothing spline functions fit to biomass concentration versus time data for the fed-batch fermentation data of Solomon, et al. (1983). The dashed line represents spline function fit using method I. The solid line represents the spline function fitted using method II.

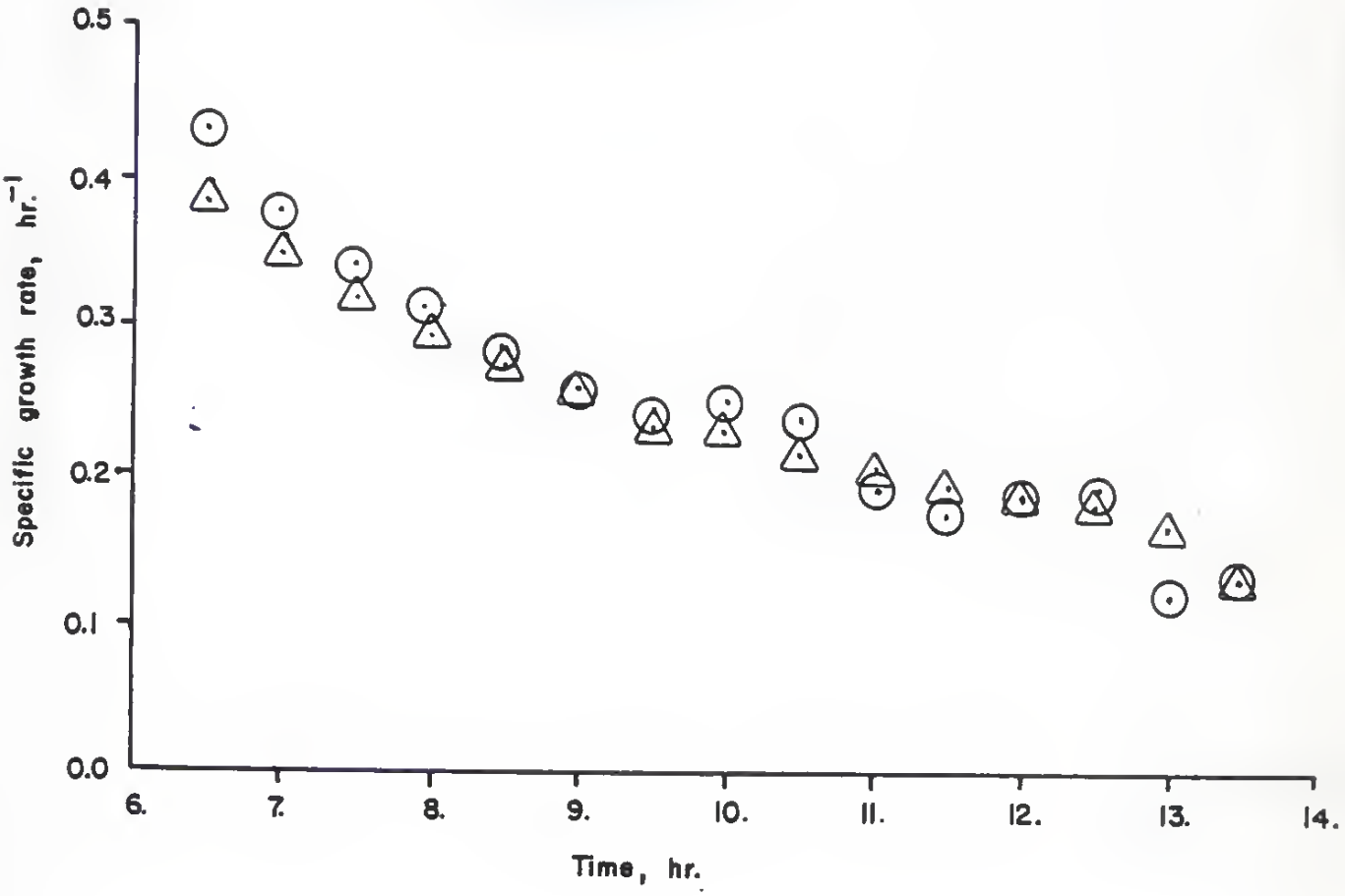


Figure 4.10 Estimates for the specific rate of microbial growth for the fed-batch data of Solomon, et al. (1983). The triangles represent estimates obtained using method I. The circles represent estimates obtained using method II.

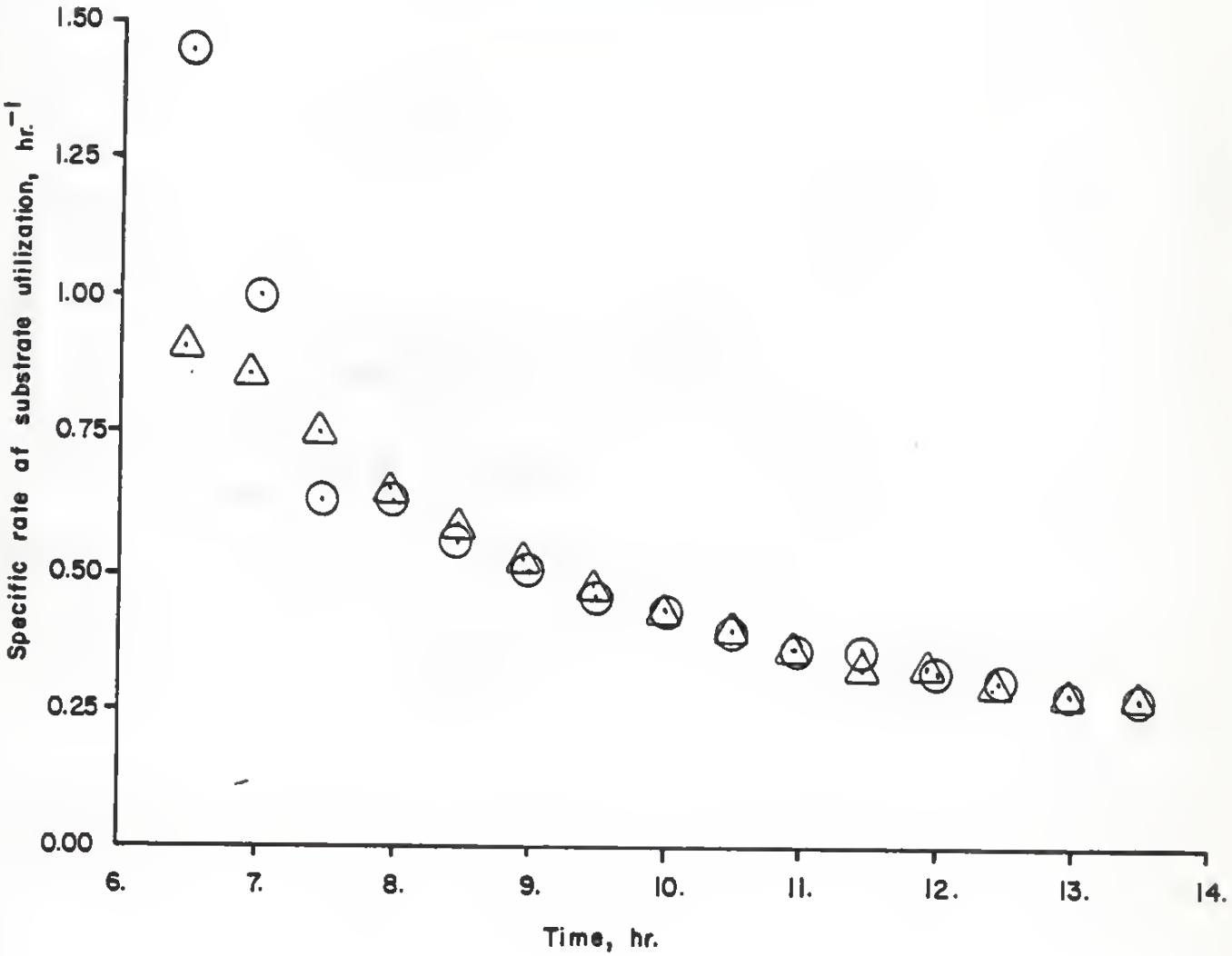


Figure 4.11 Estimates for the specific rate of substrate utilization for the fed-batch data of Solomon, et al. (1983). The triangles represent estimates obtained using method I. The circles represent estimates obtained using method II.

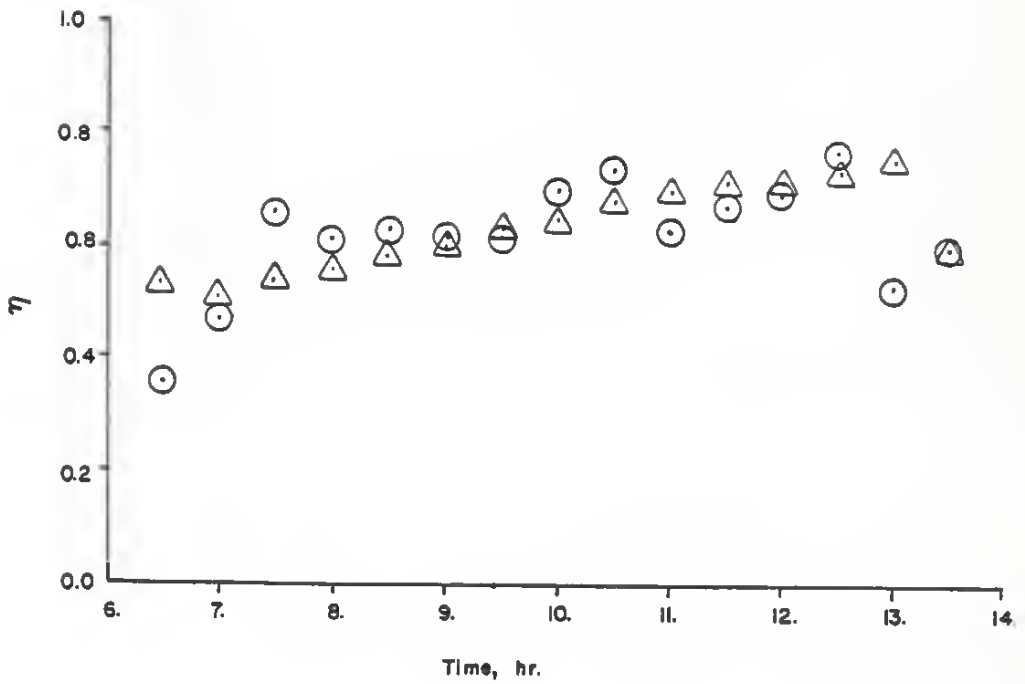


Figure 4.12 Estimates for the biomass energetic yield, for the fed-batch data of Solomon, et al. (1983). The triangles represent estimates obtained using method I. The circles represent estimates obtained using method II.

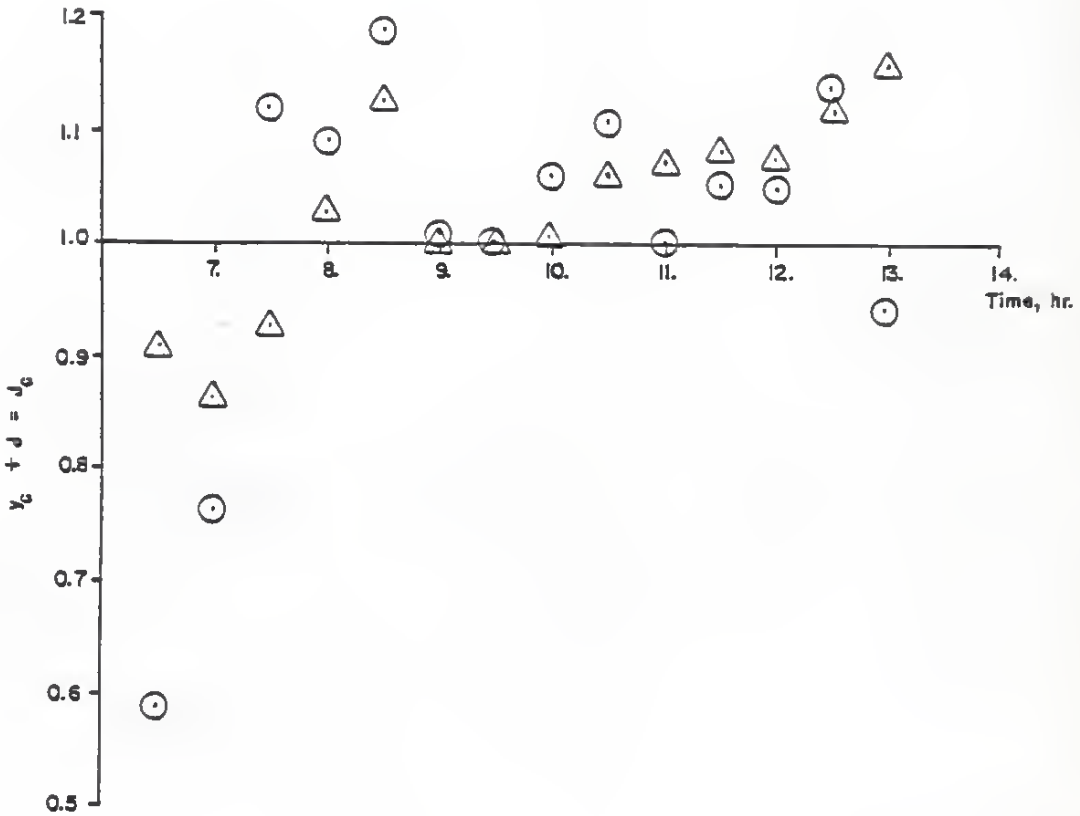


Figura 4.13 Plot of the cerbon belanca versus time for the fed-batch data of Solomon, et al. (1983). The triangles represent the balance celculated from specific rata aatimates mada using method I, and the circles rapresent the belanca celculeted from specific rate astimates made uaing method II.

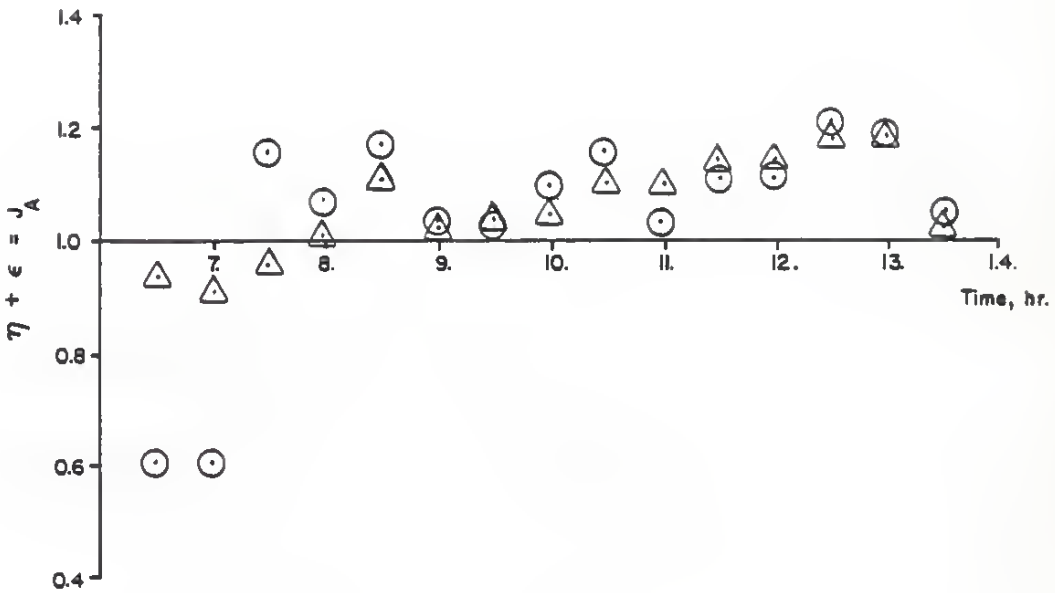


Figure 4.14 Plot of the available electron balance veraus time calculated from tha data of Solomon, et al. (1983), datermined from specific rate estimates made from the spline functions fitted to all of the data. The trianglas represent the balance calculated from specific rate estimates made using method I, and the circles represent the balance calculated from specific rate estimates made using method II.

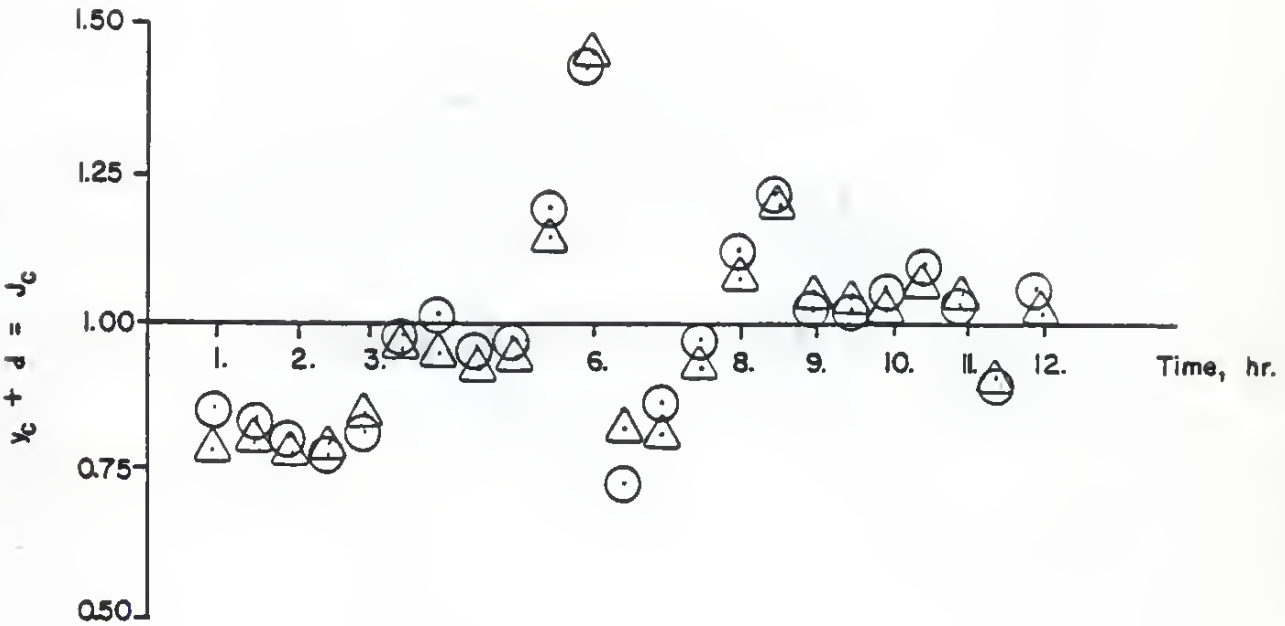


Figure 4.15 Plot of the carbon balance versus time calculated from the data of Solomon, et al. (1983), determined from specific rate estimates made from the spline functions fitted to all of the data. The triangles represent the balance calculated from specific rate estimates made using method I, and the circles represent the balance calculated from specific rate estimates made using method II.

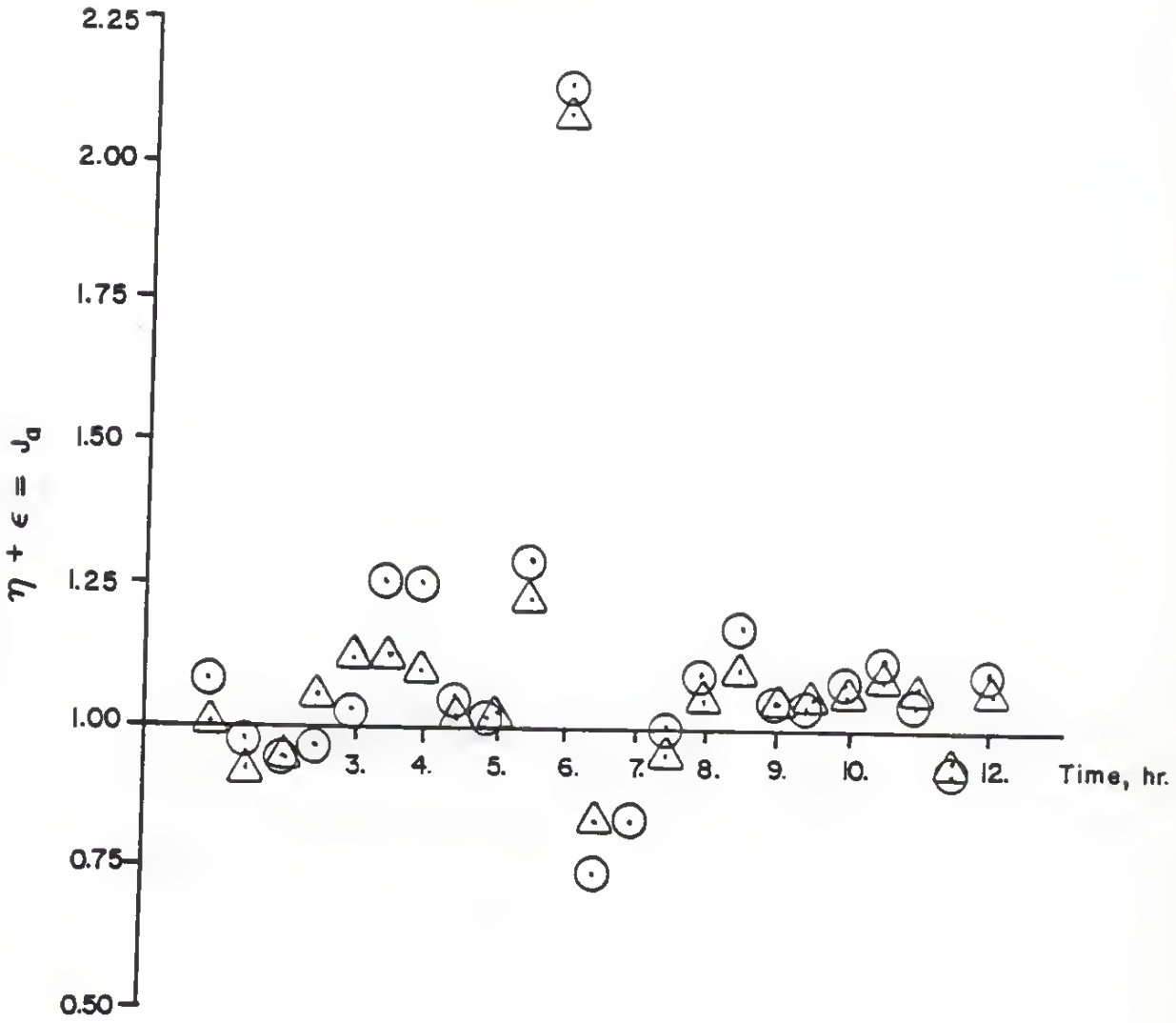


Figure 4.16 Plot of the available electron balance versus time, calculated for the data of Solomon, et al. (1983), determined from specific rate estimates made from the spline functions fitted to all of the data. The triangles represent the balance calculated from specific rate estimates made using method I, and the circles represent the balance calculated from specific rate estimates made using method II.

Table 4.1. Estimates of the True Biomass Energetic Yield and Maintenance Coefficient Determined Using Three Procedures for Calculation of the Specific Growth Rate and Specific Rate of Substrate Utilization for the Case Where the Batch and Fed-Batch Data were Smoothed Separately.

		True Growth Yield, η_{max}					
Data	Covariates Included	Point(S)	Point(RS)	Point(CV)	CI(S*)	CI(RS)	CI(CV)
Q_s, μ	none	0.567	0.652	0.609	[0.492, 0.668]	[0.595, 0.723]	[0.510, 0.757]
Q_o, μ	none	0.532	0.564	0.549	[0.494, 0.577]	[0.536, 0.596]	[0.506, 0.601]
Q_c, μ	none	0.653	0.756	0.714	[0.607, 0.708]	[0.695, 0.830]	[0.658, 0.781]
All	Z_1, Z_2	0.612	0.649	0.617	[0.548, 0.694]	[0.605, 0.699]	[0.556, 0.693]
All	Z_1	**0.603	**0.594	**0.585	[0.563, 0.649]	[0.549, 0.646]	[0.541, 0.637]
All	none	0.580	0.591	0.703	[0.532, 0.637]	[0.494, 0.727]	[0.609, 0.832]

		Maintenance Coefficient, m_e, $hr.^{-1}$					
Data	Covariates Included	Point(S)	Point(RS)	Point(CV)	CI(S)	CI(RS)	CI(CV)
Q_s,	none	-0.029	0.029	0.018	[-0.100, 0.042]	[-0.033, 0.091]	[-0.117, 0.153]
Q_c,	none	-0.034	-0.008	-0.012	[-0.073, 0.005]	[-0.047, 0.031]	[-0.079, 0.144]
Q_c,	none	0.020	0.085	0.065	[-0.011, 0.051]	[-0.014, 0.134]	[0.015, 0.115]
All	Z_1, Z_2	0.006	0.035	**0.023	[-0.032, 0.044]	[-0.012, 0.082]	[-0.052, 0.098]
All	Z_1	**0.002	**0.013	0.008	[-0.028, 0.033]	[-0.030, 0.056]	[-0.048, 0.064]
All	none	-0.014	0.011	0.061	[-0.055, 0.027]	[-0.066, 0.088]	[0.002, 0.120]

*S = Solomon et al.[1]; RS = Response Surface; CV = Cross Validation; CI = 95% Confidence Internal.
**Results with shortest confidence interval for which point estimate of m_e is positive.

Table 4.2. Estimates of the True Biomass Energetic Yield and Maintenance Coefficient Determined Using Three Procedures for Calculation of the Specific Growth Rate and the Specific Rate of Substrate Utilization for the Case Where the Batch and Fed-Batch Data were Smoothed as One Data Set.

		True Growth Yield, η_{max}					
Data	Covariates Included	Point(S)	Point(RS)	Point(CV)	CI(S*)	CI(RS)	CI(CV)
Q_s, μ	none	0.567	0.498	0.450	[0.492, 0.668]	[0.459, 0.544]	[0.388, 0.535]
Q_o, μ	none	0.532	0.547	0.479	[0.494, 0.577]	[0.501, 0.602]	[0.442, 0.522]
Q_c, μ	none	0.653	0.699	**0.637	[0.607, 0.708]	[0.648, 0.758]	[0.587, 0.697]
All	Z_1, Z_2	0.612	0.724	0.637	[0.548, 0.694]	[0.584, 0.953]	[0.536, 0.786]
All	Z_1	**0.603	0.569	0.529	[0.563, 0.649]	[0.527, 0.619]	[0.489, 0.575]
All	none	0.580	**0.570	0.510	[0.532, 0.637]	[0.531, 0.614]	[0.464, 0.568]

		Maintenance Coefficient, m_e, $hr.^{-1}$					
Data	Covariates Included	Point(S)	Point(RS)	Point(CV)	CI(S)	CI(RS)	CI(CV)
Q_s,	none	-0.029	-0.050	-0.137	[-0.100, 0.042]	[-0.129, 0.029]	[-0.292, 0.030]
Q_o,	none	-0.034	0.051	-0.059	[-0.073, 0.005]	[-0.027, 0.129]	[-0.136, 0.019]
Q_c,	none	0.020	0.110	**0.051	[-0.011, 0.051]	[0.057, 0.162]	[-0.010, 0.112]
All	Z_1, Z_2	0.006	0.117	0.059	[-0.032, 0.044]	[0.027, 0.207]	[-0.024, 0.144]
All	Z_1	**0.002	0.036	-0.048	[-0.028, 0.033]	[-0.023, 0.095]	[-0.079, 0.059]
All	none	-0.014	**0.037	-0.048	[-0.055, 0.027]	[-0.022, 0.096]	[-0.138, 0.041]

*S = Solomon et al.[1]; RS = Response Surface; CV = Cross Validation; CI = 95% Confidence Interval.
**Results with shortest confidence interval for which point estimate of m_e is positive.

Table 4.3 Consistency of Specific Rate Estimates for the 5-3-82 Batch/ Fed-Batch Data of Solomon, et al.[1].

Portion of Data Smoothed	Method	F Value
Batch	I	0.0201
Batch	II	0.0966
Fed-Batch	I	0.0166
Fed-Batch	II	0.0472
All	I	0.0857
All	II	0.1001

CHAPTER V

ANALYSIS OF EXPONENTIAL GROWTH: ANALYSIS OF YIELD, TIME INTERVAL, AND MAXIMUM SPECIFIC GROWTH RATE

INTRODUCTION

The batch fermentation is used for experimental studies and also for production purposes. The methods of data analysis for batch fermentation are difficult because the specific growth rate and the specific rate of substrate consumption are derivative quantities. The growth yield and otheryield parameters may change during the course of the fermentation due to changes in the specific growth rate and environmental conditions. During exponential growth, the specific rates and yields should remain relatively constant. Data analysis may be carried out to identify the interval of exponential growth and estimate the parameter values. While some work has been reported,[1,2] additional theoretical development and results are presented in this chapter. A method to select the exponential growth interval is described. The maximum specific growth rate is estimated from biomass, substrate, oxygen, and carbon dioxide data. Yield parameters are estimated assuming that the yield remains constant during exponential growth. The experimental results of Solomon et al.[1] are used to illustrate the methods.

THEORY

Consider the case of aerobic microbial growth with glucose as organic substrate and ammonia as nitrogen source; that is,

$$CH_mO_\ell + aNH_3 + bO_2 = y_cCH_pO_nN_q + cH_2O + dCO_2 \tag{5.1}$$

where the carbon balance is

$$y_c + d = 1.0 \tag{5.2}$$

and the available electron balance is[1]

$$\varepsilon + \eta = 1.0 \tag{5.3}$$

where ε is the fraction of available electrons transferred to oxygen and η is the fraction of available electrons incorporated into microbial biomass.

During the interval of exponential growth, the yield parameters should be approximately constant. Linear regression may be used to estimate the yield parameters, y_c, η, ε, and d. Substrate and biomass data may be used to estimate y_c and η; that is,

$$X = \frac{\sigma_s \gamma_s}{\sigma_b \gamma_b} \eta \, (S_o - S) + X_o \tag{5.4}$$

and

$$y_c = \frac{\gamma_s \eta}{\gamma_b} \tag{5.5}$$

Oxygen and substrate measurements may be used to estimate ε using the expression

$$\int_{t_o}^{t} Q_o \, X \, dt = \frac{\sigma_s \gamma_s \varepsilon}{48} (S_o - S) \tag{5.6}$$

The biomass yield based on oxygen consumption η/ε may be estimated directly from oxygen and biomass data; that is,

$$X = \frac{48\eta}{\sigma_b \gamma_b \varepsilon} \int_{t_o}^{t} Q_o \, X \, dt + X_o \tag{5.7}$$

Carbon dioxide and substrate data may be used with the equation

$$\int_{t_o}^{t} Q_c \, X \, dt = \frac{\sigma_s d}{12} (S_o - S) \tag{5.8}$$

to estimate the fraction of carbon evolved as carbon dioxide, d. The carbon yield based on carbon dioxide production, y_c/d, may be estimated using linear regression and the biomass and CO_2 data; the appropriate equation is

$$X = \frac{12\ y_c}{\sigma_b d} \int_{t_o}^{t} Q_c\ X\ dt + X_o \qquad (5.9)$$

The specific growth rate is assumed to be constant during the interval of exponential growth. The value of the maximum specific growth rate may be estimated using all of the experimental data for the interval and the covariate adjustment method[1,2,3,4,5]. For the cell concentration data,

$$\mu = \frac{1}{X}\frac{dX}{dt} \qquad (5.10)$$

and

$$\ell nX = \mu\Theta + \ell nX_o \qquad (5.11)$$

where X_o is the cell concentration at time $t = t_o$ where exponential growth begins, and $\Theta = t - t_o$. If the estimated values of yield for the interval of exponential growth, η_1, y_{c1}, ε_1, and d_1, are used, expressions similar to eqs. (5.10) and (5.11) may be obtained for the substrate, oxygen and carbon dioxide data. Substituting Equation (5.4) and the derivative expression

$$\frac{dX}{dt} = -\frac{\sigma_s \gamma_s \eta_1}{\sigma_b \gamma_b}\frac{dS}{dt} \qquad (5.12)$$

into equation (10) gives

$$\mu = \frac{-\frac{\sigma_s \gamma_s \eta_1}{\sigma_b \gamma_b}\frac{dS}{dt}}{\frac{\sigma_s \gamma_s \eta_1}{\sigma_b \gamma_b}(S_o - S) + X_o} = \frac{1}{Z}\frac{dZ}{dt} \qquad (5.13)$$

where Z is an estimate of X based on organic substrate measurements and the estimated biomass yield, η_1, for the exponential region; that is,

$$Z = \frac{\sigma_s \gamma_s}{\sigma_b \gamma_b} \eta_1 (S_o - S) + X_o \tag{5.14}$$

Equation (5.7) may be used with the estimated yields η_1 and ε_1 to obtain an estimate of X based on oxygen data; that is,

$$V = \frac{48\eta_1}{\sigma_b \gamma_b \varepsilon_1} \int_{t_o}^{t} Q_o X \, dt + X_o \tag{5.15}$$

where V is an estimate of X. Thus,

$$\mu = \frac{1}{V} \frac{dV}{dt} \tag{5.16}$$

may be used to estimate μ based on oxygen consumption measurements and the estimated values of the yield parameters.

Equation (5.9) and the estimated yields, y_{c1} and d_1, may be used to provide an estimate of X based on carbon dioxide measurements; that is,

$$W = \frac{12 y_{c1}}{b d_1} \int_{t_o}^{t} Q_c X \, dt + X_o \tag{5.17}$$

Thus,

$$\mu = \frac{1}{W} \frac{dW}{dt} \tag{5.18}$$

where W is an estimate of X based on CO_2 data.

Equations (5.10), (5.13), (5.16), and (5.18) may be used to estimate the maximum specific growth rate, μ_{max}, associated with the exponential region of growth. Each expression may be used individually, and all of the data may be analyzed simultaneously using the covariate adjustment method.[1-5]

Equations (5.10), (5.13), (5.16) and (5.18) may be written in integrated form, respectively,

$$\ell n\ X = \mu\Theta + \ell n\ X_o \tag{5.19}$$

$$\ell n\ V = \mu\Theta + \ell n\ V_o \tag{5.20}$$

$$\ell n\ W = \mu\Theta + \ell n\ W_o \tag{5.21}$$

$$\ell n\ Z = \mu\Theta + \ell n\ Z_o \tag{5.22}$$

where $\Theta = t - t_o$ and $V_o = W_o = Z_o = X_o$. When covariate adjustment is used, the average value for the measurements is

$$\bar{Y}_{oi} = \frac{1}{4} (\ell n\ X_i + \ell n\ V_i + \ell n\ W_i + \ell n\ Z_i) \tag{5.23}$$

and minimize the error in the expression

$$\bar{Y}_{oi} = \mu\Theta_i + \sum_{k=1}^{q} d_k\ Z_{ki} + \text{error} \tag{5.24}$$

is minimized where Z_{1i}, Z_{2i}, ... Z_{qi} $(1 \leq q \leq 3)$ are a set of covariates which are linear combinations of $\ell n\ X_i$, $\ell n\ V_i$, $\ell n\ W_i$ and $\ell n\ Z_i$ and have zero expected value. An example of a full set of linearly independent covariates that have zero expected value is

$$Z_{1i} = -3\ \ell n\ X_i - \ell n\ Z_i + \ell n\ W_i + 3\ \ell n\ V_i \tag{5.25}$$

$$Z_{2i} = \ell n\ X_i - \ell n\ Z_i - \ell n\ W_i + \ell n\ V_i \tag{5.26}$$

$$Z_{3i} = -\ell n\ X_i + 3\ell n\ Z_i - 3\ell n\ W_i + \ell n\ V_i \tag{5.27}$$

The slope, μ, which is estimated for the exponential region is an estimate of the maximum specific growth rate, μ_{max}.

Graphical analysis of $\ell n\ X$ vs t may be used to identify an approximate time interval of exponential growth. Statistical methods can be used with equations of the form

$$\ln X = A_1 + B_1 t + C_1 t^2 \tag{5.28}$$

$$\ln V = A_2 + B_2 t + C_2 t^2 \tag{5.29}$$

$$\ln W = A_3 + B_3 t + C_3 t^2 \tag{5.30}$$

$$\ln Z = A_4 + B_4 t + C_4 t^2 \tag{5.31}$$

to select the time interval of exponential growth. One set of points at a time may be added to the time interval of exponential growth; the statistical parameters

T = quadratic term test statistic

MSE = root mean square error

R = absolute value of residual of the added data point

can be examined to evaluate the desirability of including the added point in the time interval of exponential growth.

Equations (5.28) - (5.31) were used with SAS.[6] The observed significance level of the test statistic, T, determines whether the quadratic term is needed to fit the data to the model. For small values of the root mean square error, MSE, are desired. The residual associated with the added point should be small.

The exponential growth interval is selected by adding one set of data points at a time and examining the values of T, MSE , and R which are provided in SAS.[6]

Pirt's model[7] may be written in the form

$$\frac{\mu}{\eta} = \frac{\mu}{\eta_{max}} + m_e \tag{5.32}$$

The available electron balance and carbon balance may be used with Eq. (5.32) to obtain

$$\frac{\mu(\varepsilon + \eta)}{\eta} = \frac{\mu}{\eta_{max}} + m_e \tag{5.33}$$

and

$$\frac{\mu(y_c + d)}{\eta} = \frac{\mu}{\eta_{max}} + m_e \tag{5.34}$$

Point and interval estimates of η_{max} and m_e may be obtained using each of the above three equations individually. Covariate adjustment[1,3,4,5] may be used to estimate η_{max} and m_e using all of the measurements, simultaneously.

The data used to estimate the true growth yield, η_{max}, and the maintenance coefficient, m_e, include the selected point estimates of μ_{max}, η_1, ε_1, y_{c1}, and d_1 from smoothing the fed-batch data using spline functions. The process of smoothing the data is described in Chapters 2, 3, and 4.

MATERIALS AND METHODS

The data set examined was the 5-3-82 data set of Solomon *et al.*[1] Integration of the oxygen and carbon dioxide data was performed with the trapezoidal rule with $\Delta t = 0.5$ hrs. The time interval from 1.5 to 5.0 hrs. was selected as the initial interval from the plot of ℓn X vs t. All regressions were carried out with SAS.[6]

RESULTS AND DISCUSSION

A plot of ℓn X vs time for the batch data of Solomon *et al.*[1] is shown in Figure 5.1. The region of the curve which appears most linear is from

t = 1.5 to 5 = 5.0 hr. Regression of Eq. (5.4) for this time interval gave η_1 = 0.6186 with a 95% confidence interval of (0.5384, 0.6988). Similarly, regression of Eqs. (5.6) and (5.8) gave ε_1 = 0.5027 with (0.4439, 0.5615) as the 95% confidence interval and d_1 = 0.3553 with (0.3309, 0.3797) as the 95% confidence interval estimate.

Starting with the interval from t = 1.5 to 5.0 hrs, sets of data points were added to one end of the interval at a time to obtain the results in Table 5.1 for the cell concentration data, and Table 5.2 for the other three measured values. Analysis of the results in Tables 5.1 and 5.2 shows that the time interval 1.5 to 5.5 has consistently small values of MSE and R. The values of T are larger than 0.1 for this time interval except for V; however, values of T were small for all values of V that were tested. The values of R for the interval 1.5 to 5.5 are those associated with 5.5 hrs.

Point and 95% confidence interval estimates of the maximum specific growth rate, μ_{max}, and the length of the confidence interval are reported in Table 5.3. Covariate adjustment is used with several different sets of covariates. The shortest 95% confidence interval is found when two covariates, Eqs. (5.25) and (5.27), are used in Eq. (5.24). The 95% confidence interval from 0.482 to 0.506 includes all of the point estimates except for μ_{max} = 0.470 which is obtained using the oxygen data. Thus, there is good agreement when the different estimates are compared.

Table 5.4 contains the results of regression analysis using Eqs. (5.32), (5.33), and (5.34), the covariate adjustment method, the results of analysis of the exponential region, and the results of smoothing the fed-batch data. The reciprocal of the true growth yield and the maintenance parameter are estimated using each data set individually and all data simultaneously. The covariates

$$Z_{1i} = -\frac{\mu_i}{\eta_i} + \frac{\mu_i(y_{ci} + d_i)}{\eta_i} \tag{5.35}$$

$$Z_{2i} = \frac{\mu_i}{\eta_i} - \frac{2\mu_i(\eta_i + \varepsilon_i)}{\eta_i} + \frac{\mu_i(y_{ci} + d_i)}{\eta_i} \tag{5.36}$$

were used with the covariate adjustment method.

The exponential growth data point is $\mu_{max} = 0.494$, $\eta_1 = 0.6186$, $\varepsilon_1 = 0.5027$, $d_1 = 0.3553$, and $y_{c1} = 0.5766$. The other data which are the result of smoothing the fed-batch data using spline functions are presented in Chapter 4.

The first and last data points for the batch and fed-batch growth intervals were deleted because the estimated derivatives are one sided. The results of data smoothing gave an estimate of specific growth rate at t = 1.0 hr which is considerably larger than the estimate for the exponential interval; thus, the point at t = 1.0 hr was not included in estimating the yield parameters.

The results in Table 5.4 show that the maintenance parameter is small and negative in all cases except one. The shortest 95% confidence interval for η_{max} is 0.076 and appears twice (for the oxygen data and for all data with no covariates). The shortest 95% confidence interval for m_e occurs when one covariate is used; the next shortest interval occurs for all data with no covariates. The estimates with no covariates and one covariate are similar. These estimates are also in reasonable agreement with the estimates obtained previously as shown in Table 5.5. The estimate of growth yield for the exponential region, $\eta_1 = 0.6186$, is also in reasonable agreement with the estimates of true growth yield.

CONCLUSIONS

A systematic procedure to analyze the exponential growth region of a batch fermentation is presented. Point and interval estimates of the yield

parameters and the maximum specific growth rate are estimated. The covariate adjustment method was used together with biomass, substrate, oxygen, and carbon dioxide measurements to estimate the maximum specific growth rate. Statistical methods were presented snd used to select the time interval of exponential growth.

The procedure, which is presented, is easily carried out using SAS.[6] The estimated values of maximum specific growth rate and yields are used with smoothed fed-batch data to estimate the true growth yield and maintenance parameter. The results are in good agreement with earlier estimates.

REFERENCES

1. Solomon, B. O., L. E. Erickson, and S. S. Yang, "Utilization of Statistics and Experimental Design in Data Collection and Analysis," Biotechnol. Bioeng., 25, 2683 (1983).

2. Oner, M. D., L. E. Erickson, and S. S. Yang, "Yield, Maintenance and Product Formation on Anaerobic Fermentations," presented in session 39 Fundamentals of Anaerobic Fermentations at the AIChE, 1984 San Francisco Meeting: preprint 39a.

3. Solomon, B. O., L. E. Erickson, J. L. Hess, and S. S. Yang, "Maximum Likelihood Estimation of Growth Yields," Biotechnol. Bioeng., 24, 633 (1982).

4. Solomon, B. O., M. D. Oner, L. E. Erickson, and S. S. Yang, "Estimation of Parameters where Dependent Observations are Related by Equality Constraints," AIChE Los Angeles Meeting (Nov., 1982); AIChE Journal, 30, 747 (1984).

5. Yang, S. S., B. O. Solomon, M. D. Oner, and L. E. Erickson, "A Method of Estimation and Testing Common Parameters for Some Multi-Response Models Associated with Growth and Bioenergetics," Technometrics, 26, 355 (1984).

6. SAS User's Guide, SAS Institute, Raleigh, N.C. (1979).

7. Pirt, S. J., Principles of Microbe and Cell Cultivation, Blackwell, London (1975).

NOMENCLATURE

Roman Letters

A,B,C = coefficients of regression

d_1 = fraction of carbon substrate evolved as carbon dioxide during exponential growth.

m_e = maintenance coefficient, $hr.^{-1}$

MSE = square root of the mean of the square of the error of the estimate

Q_c = specific rate of carbon dioxide evolution, moles CO_2 gr biomass^{-1} $hr.^{-1}$

Q_o = specific rate of oxygen uptake, moles O_2 gr biomass^{-1} $hr.^{-1}$

R = absolute value of the residual for the point which is added

S = concentration of substrate in the fermentor, g L^{-1}

S_o = concentration of substrate in fermentor inlet, g L^{-1}

t = time

t_o = time at start of exponential interval, hr.

T = level of significance of test statistic for quadratic term in the growth curve

V = variable defined by Eq. (5.15), gr biomass L^{-1}

W = variable defined by Eq. (5.17), gr biomass L^{-1}

X = biomass concentration, gr L^{-1}

X_o = biomass concentration at start of exponential growth interval, gr L^{-1}

$\bar{Y}$ = variable defined by Eq. (5.23)

Y_{cl} = biomass carbon yield during exponential growth

Z = variable defined by Eq. (14), gr biomass L^{-1}

Z_{1i}, Z_{2i}, Z_{3i} = covariates defined by Eqs. (5.25), (5.26) and (5.27)

Greek Letters

γ_b = reductance degree of biomass, equiv. of available electrons in biomass gr atom carbon^{-1}

γ_s = reductance degree of substrate, equiv. of available electrons in biomass gr atom carbon^{-1}

ε_1 = equvalents of available elections of substrate transferred to oxygen per equivalent of substrate during exponential growth

η = biomass energetic yield

η_1 = biomass energetic yield during exponential growth

η_{max} = true growth yield

θ = $t - t_o$, hr

η = specific growth rate, hr^{-1}

μ_{max} = maximum specific growth rate, hr^{-1}

σ_b = weight fraction of carbon in biomass

σ_s = weight fraction of carbon in substrate

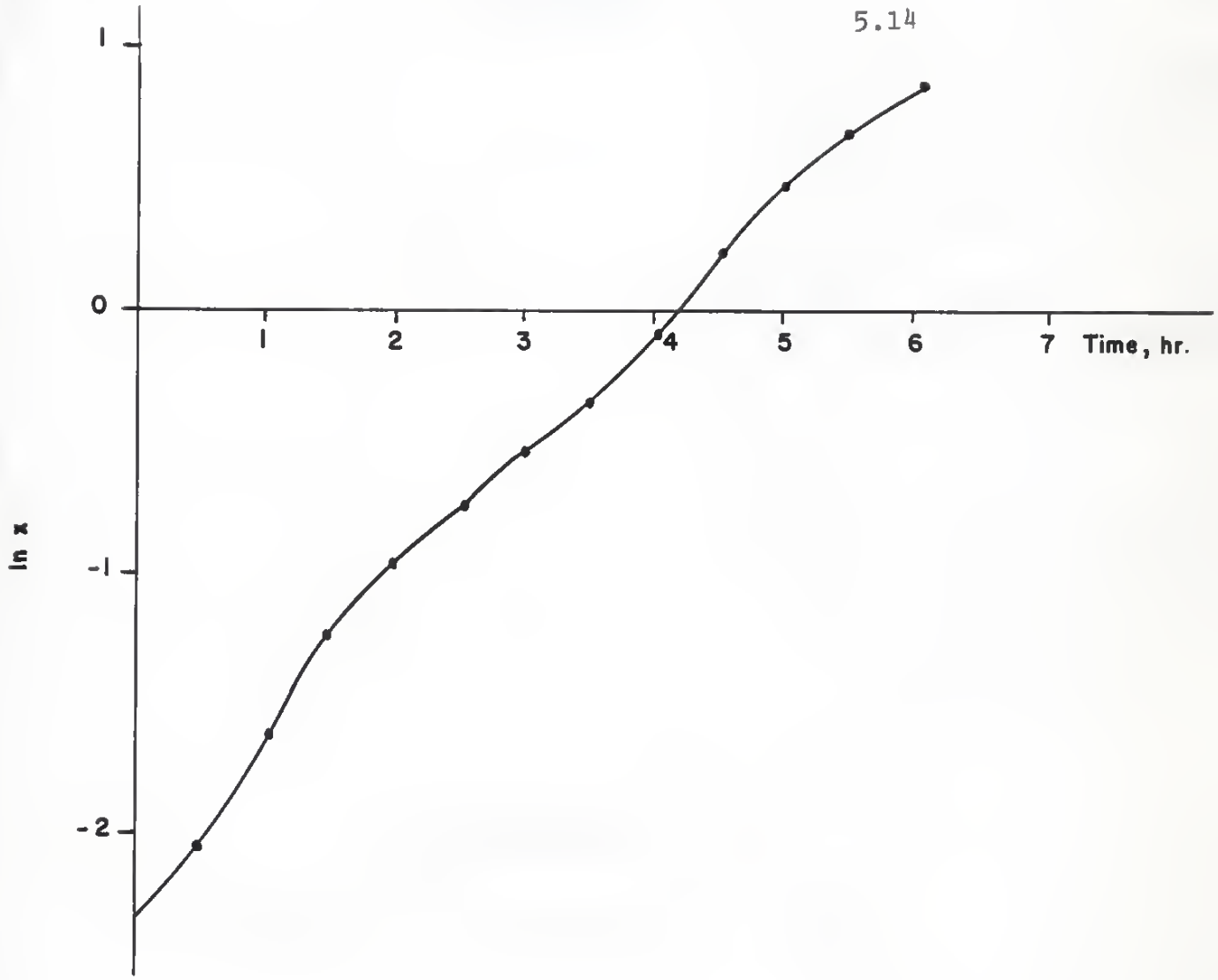


Figure 5.1 Plot of the natural logarithm of biomass concentration versus time for the batch data of Solomon.[1]

Table 5.1. Results of the regression of Eq. (5.28) for several time intervals.

Time interval	μ_{max}, hr^{-1}	T	MSE	R
0.0 - 5.0	0.5395	0.0260	0.0936	0.10645
0.5 - 5.0	0.5253	0.0964	0.0881	0.16378
1.0 - 5.0	0.4980	0.7382	0.0550	0.06020
1.5 - 5.0	0.4851	0.4589	0.0506	-------
*1.5 - 5.5	0.4815	0.7540	0.0475	0.01662
1.5 - 6.0	0.4747	0.7067	0.0478	0.04084
1.0 - 5.5	0.4927	0.4796	0.0533	0.06724
1.0 - 6.0	0.4852	0.1912	0.0551	0.05655

* Interval selected as time interval of exponential growth.

Table 5.2. Determination of the Region of Exponential Growth Using Estimates of Biomass Concentration Based on Substrate Concentration, Rate of Carbon Dioxide Evolution and Rate of Oxygen Uptake Data.

Time Interval	Z				W				V			
	μ_{max}	T	MSE	R	μ_{max}	T	MSE	R	μ_{max}	T	MSE	R
*1.5 - 5.5	0.4696	0.2182	0.0652	0.0197	0.4906	0.4033	0.0215	0.0081	0.4843	0.0001	0.0383	0.0512
1.0 - 5.5	0.4856	0.0466	0.0741	0.0959	0.5023	0.0418	0.0365	0.0700	0.4992	0.0001	0.0529	0.0890
0.0 - 5.0	0.5572	0.0004	0.1451	0.2575	0.5517	0.0003	0.1052	0.2160	0.5681	0.0001	0.1236	0.2503
0.5 - 5.0	0.5229	0.0049	0.1013	0.1825	0.5229	0.0013	0.0625	0.1163	0.5348	0.0001	0.0756	0.1426
1.0 - 5.0	0.5233	0.003	0.1012	0.0514	0.5240	0.0005	0.0591	0.0445	0.5354	0.0001	0.0713	0.0958
1.5 - 5.0	0.4976	0.0361	0.0747	------	0.5036	0.0068	0.0363	------	0.5111	0.0001	0.0428	------
1.0 - 6.0	0.4767	0.0196	0.1749	0.0671	0.4983	0.0318	0.0355	0.0222	0.4910	0.0001	0.0556	0.0612
1.5 - 6.0	0.4621	0.1213	0.0641	0.1231	0.4894	0.3376	0.0201	0.0350	0.4775	0.0001	0.0400	0.0601

*Time interval selected as exponential growth region.

Table 5.3. Point and 95% Confidence Interval Estimates for μ_{max}.

Regression Equation	Covariates Included	Point Estimate	95% Confidence Interval	Interval Length
(5.19)	none	0.482	[0.452, 0.511]	0.0598
(5.20)	none	0.470	[0.428, 0.511]	0.0824
(5.21)	none	0.491	[0.477, 0.504]	0.0268
(5.22)	none	0.484	[0.460, 0.508]	0.0483
(5.24)	none	0.482	[0.463, 0.500]	0.0376
(5.24)	Z_1	0.483	[0.401, 0.504]	0.0426
(5.24)	Z_2	0.482	[0.463, 0.502]	0.0386
(5.24)	Z_3	0.490	[0.475, 0.505]	0.0306
(5.24)	Z_1, Z_2	0.484	[0.462, 0.506]	0.0446
(5.24)	Z_1, Z_3	*0.494	[0.482, 0.506]	0.0246
(5.24)	Z_2, Z_3	0.491	[0.473, 0.510]	0.0370
(5.24)	Z_1, Z_2, Z_3	0.500	[0.482, 0.510]	0.0284

*Estimate with shortest 95% confidence interval.

Table 5.4 Point and 95% Confidence Interval Estimates for True Growth Yield and Maintenance Coefficient Calculated from the Exponential Point Estimate and Fed-Batch Results.*

Data	Covariates Included	True Growth Yield, η_{max}		Maintenance coefficient, m_e, hr^{-1}	
		Point	CI	Point	CI
Q_s, μ	none	0.584	[0.542, 0.633]	-0.033	[-0.082, 0.016]
Q_{O2}, μ	none	0.528	[0.494, 0.570]	0.039	[-0.009, 0.087]
Q_{CO2}, μ	none	0.690	[0.633, 0.758]	-0.051	[-0.101, -0.001]
All	none	0.594	[0.558, 0.634]	-0.015	[-0.055, 0.025]
All	Z_1	0.542	[0.551, 0.638]	-0.016	[-0.025, 0.025]
All	Z_1, Z_2	0.562	[0.493, 0.654]	-0.037	-0.104, 0.030]

* Fed-batch data smoothed using response surface method and spline functions. N = 14 points.

Table 5.5 Comparison of Several Different Estimates of True Growth Yield and Maintenance Parameter for Data of Solomon et al.[1].

Method	True Growth Yield, η_{max}		Maintenance Parameter, m_e, hr^{-1}	
	Point	95% Interval	Point	95% Interval
Solomon[1]	0.603	[0.563, 0.649]	0.002	[-0.028, 0.033]
RS, A	0.570	[0.531, 0.614]	0.037	[-0.022, 0.096]
CV, A	0.637	[0.587, 0.697]	0.051	[-0.010, 0.112]
RS, B	0.594	[0.549, 0.646]	0.013	[-0.030, 0.056]
CV, B	0.585	[0.541, 0.637]	0.008	[-0.048, 0.064]
RS, C	0.594	[0.558, 0.634]	-0.015	[-0.055, 0.025]

RS refers to response surface method of smoothing data.

CV refers to cross validation method of smoothing data.

A refers to all data smoothed as one set.

B refers to separate smoothing of batch and fed-batch data.

C refers to use of exponential estimate and separately smoothed fed-batch data.

MEASUREMENT OF BIOMASS CONCENTRATION USING A MICROWAVE OVEN AND ANALYSIS OF DATA FOR ESTIMATION OF SPECIFIC RATES

by

MARK ANTHONY BUONO

B.S., Lehigh University, 1983

AN ABSTRACT OF A MASTER'S THESIS

submitted in partial fulfillment of the

requirements for the degree

MASTER OF SCIENCE

CHEMICAL ENGINEERING

Department of Chemical Engineering

KANSAS STATE UNIVERSITY
Manhattan, Kansas

1985

A microwave oven has been used to measure biomass concentration and construct a growth curve for a culture of Candida utilis ATCC 1084. Drying curve results show that when a sample is exposed to microwave radiation for 15.0 minutes, microwave and air oven dry weight results are within four milligrams of one another when dried at the "cook" power setting (approximately 630 W at full power). For biomass dry weight sample sizes of 147.74 and 14.42 milligrams, the sample weight was reduced to within 10% of the final air oven dry weight in less than 4 minutes. At the "defrost" power setting 7 minutes were required to reduce the sample weight to within 10% of the final air oven dry weight for the large sample. For the small sample, this point was not reached in 15 minutes.

A new method of selecting smoothing spline functions is presented and used to curve fit experimental fermentation data. From the curves generated, estimates for the specific growth rate, the specific rate of substrate utilization and the biomass energetic yield are made. Selection of the smoothness parameters for the time profiles of biomass and substrate concentration is based on minimization of a response surface fit to the smoothness parameters. The response modelled is the extent of closure of the carbon balance and the available electron balance. Results obtained using cross validation for selection of the smoothness parameter are also presented, and compared to the results calculated using the response surface technique. Smaller values of the sum of squares of the errors of the carbon and available electron balances are obtained using the response surface method. The results obtained from smoothing the batch and fed-batch data separately are compared with the results of smoothing all of the data as one data set. While useful results are obtained in both cases, more consistent results are obtained when the batch and fed-batch data are fit separately.

Estimates made for specific growth rate and biomass energetic yield are used with the covariate adjustment procedure to calculate point and 95% confidence interval estimates for the true biomass energetic yield, η_{max}, and the maintenance coefficient m_e. The true growth yield and the maintenance coefficient are compared for the different methods.

The maximuum specific growth rate is estimated using biomass data together with substrate, oxygen and carbon dioxide data expressed as biomass equivalents. A systematic procedure is presented for analysis of the exponential region of growth using all available measurements. Yield parameters are assumed to be constant during the exponential growth time period. Point and 95% confidence interval estimates for the biomass energetic yield, carbon dioxide yield and available electron yield are made. Pirt's model is used to calculate point and 95% confidence interval estimates for the true growth yield, $1/\eta_{max}$, and maintenance coefficient, m_e. In the model, for the exponential time period, estimates for the yield quantities for the exponential region of growth and the maximum specific growth rate are used. Estimates for specific growth rate and the biomass yield obtained using the response surface method are used in the model for time periods which are not part of the exponential growth interval.